

UNIVERSIDADE DE LISBOA



FACULDADE DE LETRAS

A Responsabilidade Legal de Veículos Autónomos:
Uma Perspetiva da Ciência Cognitiva

Beatriz Assunção Ribeiro

Tese orientada pelo Prof. Doutor Hélder Coelho, Prof. Doutor João Branquinho e Dra. Ana Elisabete Ferreira,
especialmente elaborada para a obtenção do grau de mestre em Ciência Cognitiva.

Dissertação

2018/2019

Agradecimentos

Aos destinatários naturais deste agradecimento, porque acreditam na junção das áreas como contributo importante e porque sem eles este projeto nunca seria possível: Prof. Doutor Hélder Coelho, Prof. Doutor João Branquinho e Dr.^a Ana Elisabete Ferreira.

À minha turma do Mestrado de Ciência Cognitiva, cheia de pessoas brilhantes, que acreditam no futuro e que gostam de debater o mundo com a mesma curiosidade que têm as crianças, que me instigaram a seguir esta ideia quando quase desisti dela.

À minha família, em especial, pai, mãe, irmã e avó, com quem partilhei as vitórias e frustrações e me ouviram sempre.

À Lúcia, companheira de todas as horas infinitas e que acredita em mim daqui à lua, mesmo nas noites em que a lua não aparece.

Resumo

A Tecnologia dos Veículos Autônomos avança rapidamente e urge um debate entre todas as áreas que possam oferecer um contributo. O funcionamento complexo de um Veículo Autônomo, as querelas éticas associadas ao mesmo, e os problemas de segurança que se têm vindo a colocar, tornam-no num produto em tudo excecional, o que acarreta limitações ao nível do Direito. Este é o contributo da Ciência Cognitiva, com recurso às disciplinas de Filosofia da Mente e Ciências da Computação, fazendo a ponte com o Direito, com o objetivo de responder a algumas dessas limitações.

O propósito geral desta dissertação é perceber o que é um Agente, que impacto tem no ambiente e o que é que torna um Agente uma entidade legalmente responsável. Como entidades num mundo em que muito é definido pelos atos voluntários, os Veículos Autônomos são Agentes complexos enquadrados num ambiente com infinitos estímulos. A esta tecnologia ainda lhe falta, como iremos argumentar, um pressuposto mínimo da imputação da responsabilidade: a Metacognição. O que esta dissertação pretende é mostrar, numa primeira fase e através de considerações sobre Agentividade¹ e ação voluntária, que a Metacognição é um elemento essencial, no Direito, para a atribuição de Responsabilidade. Numa segunda, alega-se que o tipo de Metacognição na qual o Direito se baseia está em falta no VA. E, finalmente, debate-se qual deverá ser o regime jurídico do VA, tendo tudo isto em conta.

Para esta discussão, além de tecermos ideias sobre um regime mais disperso e solidário de responsabilidades, bem como sobre a necessidade de um sistema legal de responsabilidade civil paralelo a um seguro de responsabilidade civil obrigatório, comparamos os VA com os regimes dos menores, animais e interditos, dada a sua semelhança no que diz respeito aos processos metacognitivos.

¹A palavra mais imediata para expressar a qualidade daquele que é um Agente talvez fosse Agência. Contudo, verificando o significado desta expressão no dicionário, chegamos à conclusão que a palavra não denota o conteúdo semântico pretendido. Por outro lado, a própria utilização da palavra está associada, no mundo corrente, a empresas ou pessoas que agem em nome de outrem, que não é o nosso objetivo. Por esta razão opta-se por utilizar a expressão Agentividade.

Abstract

As the technology of Autonomous Vehicles advances swiftly, it's urgent to look into all the areas that can somehow contribute to a useful debate. The highly complex functioning of the Autonomous Vehicle, the ethics questions associated, and the safety problems that appeared with it, make it an exceptional product, with serious implications in Law. In this dissertation we pretend to build a bridge between Philosophy of Mind and Computation Science and the Law, in order to respond to the limitations of this technology.

The purpose of this thesis is to analyse what it means to be an Agent, with an impact in the environment and what makes an Agent an entity to whom we can ascribe legal responsibility. As entities in a world where a lot of what happens it's defined by the idea of Voluntary Acts, Autonomous Vehicles are complex agents, framed in an environment that presents an infinite number of stimuli. This Technology, as we shall argue, lacks a minimum requirement of responsibility: Metacognition. What this dissertation wants to show is, using the concepts of Agency and Voluntary Action, that Metacognition is an essential element, in Law, in order to ascribe responsibility. In a second phase of this thesis, we argue that this type of Metacognition, that Law requires to consider someone responsible, is missing in the Autonomous Vehicle. Finally, in the last chapters, there's some debate over the Legal system that should be applied to this technology.

In this debate, while suggesting a scattered regime that also considers joint liabilities, as well as the existence of a mandatory insurance parallel to the civil Responsibility regime, we briefly compare Autonomous Vehicles to minors, animals and disabled people, given their similarity when it comes to the degree of Metacognitive processes.

Índice

Orientações de Leitura	11
Introdução	13
Considerações Gerais sobre os Veículos Autónomos	17
1. O seu funcionamento	17
1.1. A perceção	19
1.1.1. A perceção Interna e Ambiental	19
1.1.2. A Localização	22
1.2. O Planeamento	23
1.3. O controlo	24
2. Análise do Relatório da União Europeia sobre Veículos Autónomos	26
2.1. Relatório Final apresentado à União Europeia por Dr. E.F.D. (Esther) Engelhard and R.W. (Roeland) de Bruin, LL.M. (anexo ao <i>A Common EU Approach to Liability Rules and Insurance for Connected and Autonomous Vehicles</i>)	29
3. Breves considerações Éticas	35
Introdução à Agentividade	41
1. Agentividade na Filosofia	42
2. Agentividade na Computação	45
3. Agentividade no Direito	47
4. Conclusões	51
O Ato Voluntário como elemento comum	53
A Metacognição	59
1. A Metacognição enquanto pressuposto mínimo da imputação	59
2. A Metacognição na Inteligência Artificial	62
Discussão sobre a Responsabilidade Legal	65
Conclusão	71
Referências	73

Orientações de Leitura

Numa primeira nota, importa referir que, em momento algum, a discussão irá considerar temas como consciência ou livre arbítrio. Evitar-se-ão ao máximo, embora reconhecendo, de antemão, a dificuldade do que se propõe, conceitos que possam ser, ambíguos ou cujos limites não estejam suficientemente delineados. Se assim não fosse, ver-nos-íamos, mais cedo ou mais tarde, embrenhados entre ideias que se interligam e que se explicam umas às outras de forma circular. Sem caso-base, não encontraríamos escapatória. Até porque determinados assuntos, tais como o livre arbítrio, no caso do ser humano, são de tal modo debatidos em Filosofia, sem que se encontre consenso, que não se afiguram úteis no caso da Inteligência Artificial, porque não conseguimos confirmá-los.

Partiremos, também, do pressuposto de que os seres humanos são criaturas racionais, com capacidade de compreensão² e com possibilidades de se conformar com as proposições normativas através da sua ação intencional³, caso contrário, a lei não teria qualquer valor (Morse, 2000). Isto implica, para os Agentes responsáveis, a habilidade global de agarrar conceitos e compreender os seus significados em cada momento.

Importa referir, ainda, que todo este estudo pressupõe, uma conceção antropomórfica do ordenamento jurídico, até porque o Direito não pode ser visto à luz de outra visão. Por outras palavras, entende-se que o Direito tem como destinatários últimos o ser humano e só ele ou, em alternativa, coisas ou seres sencientes em função da natureza humana.

A utilização da expressão intencionalidade, ao longo desta tese, é ambígua, no sentido em que pode tomar a forma de duas aceções distintas. Uma relaciona-se com a vontade de praticar um ato; na restante estão em causa os estados mentais intencionais, isto é, pensamentos que versam sobre um determinado objeto (por exemplo, dizer que “A Casa Branca é em Washington” é um pensamento e uma afirmação sobre a Casa Branca e sobre um traço característico de Washington).

Finalmente, é sabido que até uma nova tecnologia seja efetivamente utilizada, os seus riscos são sempre difíceis de prever e este facto é algo que é impossível de contornar e, logo, esta dissertação deverá ser lida tendo sempre esta incerteza em mente.

Siglas:

VA - Veículo Autónomo

UE - União Europeia

IA - Inteligência Artificial

DRRDP - Diretiva Relativa à Responsabilidade Decorrente dos Produtos

EM - Estado Membro

² Excluindo, claro está, o caso dos inimputáveis.

³ Neste sentido note-se que se trata de intenção na ideia de propósito ou objetivo.

Introdução

O Direito depende, em boa parte, do processo de classificação; é a classificação de um problema, enquanto situação que pode ser enquadrada em determinada norma legal, que permite resolver uma questão jurídica (Birks, 2000); e é à conta desta ideia de categorização que o estatuto jurídico de uma entidade lhe confere certos direitos e deveres, necessária e diretamente relacionados com a sua natureza (Adams, 2010).

Deste modo, faz sentido que para Anne Boulangé e Carole Jaggie (2014), o primeiro passo para a determinação de um regime legal do Robô é determinar o estatuto do Robô, considerando as diversas opções plausíveis, a saber: (1) o estatuto de objeto, (2) pessoa jurídica equivalente ao ser humano, (3) pessoa jurídica distinta do ser humano, (4) de animal ou (5) estatuto novo e autónomo (que, a existir, pode variar conforme o tipo de Robô em causa).

Muito embora a Jurisprudência tenha vindo a fazer um trabalho no sentido de determinar um estatuto dos Veículos Autónomos, esta não será, a longo prazo, uma opção viável, porquanto se espera que utilização de Robôs aumente e, com isto, aumente em proporção o número de casos por danos causados por, eventualmente entupindo os tribunais, visto que na falta de um regime legal, as partes não sabem que Direitos têm, nem tampouco a sua extensão, e apenas deste modo podem conhecê-los. Outra causa para esta sobrecarga dos tribunais é a quantidade de particularidades inerentes à tecnologia aproveitada pelos VAs, que certamente implicariam inúmeros pedidos de interpretação da norma geral e aplicação ao caso concreto. Adicionalmente, os VAs em circulação, principalmente na ausência de legislação, têm colocado problemas gravíssimos ao nível da segurança pública (veja-se o caso da Tesla, cujo protótipo que estava a ser testado vitimou um peão, nos EUA, em março de 2018), com o quais não devemos compactuar. Estas são as razões que nos obrigam a discutir todos os conceitos inerentes à tecnologia, bem como a legislar sobre o tema.

Assim, importa ver as sugestões que têm vindo a ser feitas para a legalização desta tecnologia, fazendo sempre uso do dualismo jurídico entre coisas e pessoas (Pietrzykowski, 2017). O problema da utilização de dualismos é que nem sempre podem refletir a realidade da forma mais fiel, uma vez que só admitem dois extremos. Por exemplo, no caso dos animais, não parece haver muitas dúvidas sobre se exibem ou não formas simples de consciência (Andrews, 2014; Churchland et al., 2014) e é por isso que acabam por não ser, à luz do direito, nem objetos nem sujeitos jurídicos. Parece bastante óbvio que existe uma classe de entes que não mais faz sentido serem vistos como objetos, mas que ao mesmo tempo ainda não são considerados sujeitos, o que, a longo prazo, se pode tornar insustentável (Pietrzykowski, 2017).

Urge, portanto, o debate sobre estas ideias, olhando para todas as áreas do conhecimento e colhendo o que cada uma nos tem para nos oferecer. É neste sentido que, nesta dissertação se oferece uma perspetiva da Ciência Cognitiva, balançando entre a Filosofia da Mente, as Ciências da Computação e Inteligência Artificial, com algumas contribuições da Psicologia Cognitiva, desembocando, finalmente, no Direito que, neste momento, necessita de tanta ajuda quanto possível no estudo de sistemas de Inteligência Artificial.

Considerando o ora exposto, Ugo Pagallo (2013), resumiu a questão de forma muito pertinente, enunciando três hipóteses para uma descrição normativa da Inteligência Artificial e, para o que nos convém, de um Veículo Autônomo, conforme se verifica na figura 1. Em conjugação com as ideias de Boulangé e Maggie, acima descritas, o Veículo Autônomo poderá vir a ser entendido como (1) pessoa jurídica, dotada de direitos e deveres, em toda a extensão possível, equivalente ao ser humano, (2) uma entidade própria, com um estatuto próprio, ou (3) um mero objeto, apenas fonte de dano.

Responsible robot	<i>Immunity</i>	<i>Strict liability</i>	<i>Unjust damages</i>
As legal person	I-1	SL-1	UD-1
As proper agent	I-2	SL-2	UD-2
As source of damage	I-3	SL-3	UD-3

Fig. 1 - Os nove tipos ideais de responsabilidade, de acordo com Ugo Pagallo.

Este quadro, da figura 1, ajudar-nos-á a manter em mente que, se não o objetivo, então a consequência última, será a localização do VA, enquanto entidade, neste mesmo quadro.

Nas categorias acima mencionadas, o que importa, numa primeira análise, é perceber o que significa ser um Agente. Se uma entidade é um Agente, não poderá ser uma coisa, dado que a Lei da não contradição não o permite. Atendendo ao facto que uma coisa se opõe à outra (e, de facto, opõem-se, visto que apresentam características opostas) a frase *O Robô é um Agente* e a frase *O Robô é uma coisa* não podem, nunca, ser as duas verdadeiras ao mesmo tempo. Por exemplo, um Agente, como veremos, age voluntariamente, enquanto uma coisa não age de todo. É lógico que uma entidade não pode agir voluntariamente e não agir de todo simultaneamente.

Para tal, salvo considerações fortuitas sobre outros temas, **iremos debruçar-nos sobre o conceito de Agentividade e, mais importante, sobre o que significa ser um Agente, quer no domínio da Filosofia quer no domínio da Inteligência Artificial e do Direito, porque o conceito de personalidade jurídica anda de mãos dadas com a ideia de Agentividade** (Asaro, 2007). Só depois de compreendermos o que é um Agente, é que nos podemos questionar sobre o que faz de um Agente uma entidade à qual podemos atribuir responsabilidade legal.

Porque viemos a ser confrontados com o facto de o conceito de Agentividade ser, de todo o modo, árduo de tratar do ponto de vista de uma uniformização nas três áreas em causa, conforme, aliás, já Shardlow (1995) havia afirmado, vimo-nos obrigados a tentar descobrir um ponto transversal nas teorias das quais nos estávamos a servir para compreender o tema. Só encontrando um denominador comum nas três áreas, é que poderíamos entender qual deveria ser o pressuposto da Agentividade.

Foi então que notámos que, independentemente da teoria sobre Agentividade que utilizemos para defender certa visão sobre o estatuto do VA, o conceito de Ato Voluntário está, inevitavelmente, presente. Um Agente é-o na medida em que age voluntariamente, sem prejuízo de condições adicionais que queiramos acrescentar. Se uma entidade é dotada da capacidade de agir voluntariamente, então, é um Agente, e, consequentemente, pode ser responsabilizado, já nos dizia Aristóteles em *Ética a Nicómaco*.

Surge, assim, a questão de saber por que razão não se pune, em Direito, as crianças e os animais, em nome de um Ato Voluntário, se é esta a chave que abre a porta para o significado de Agente. Neste ponto, sim, indagamos sobre o que torna um Agente uma entidade responsável. A resposta, se argumentará, está na Metacognição, conceito amplamente discutido na Filosofia e na Psicologia e do qual o Direito se serviu durante séculos sem que um nome lhe fosse atribuído. Demos conta, então, que existe um conceito, em Direito, que, não obstante *anónimo*, tem um papel fulcral na vida corrente do processo legal.

Deste modo se concluiu que é um Agente o ente que age voluntariamente, mas que nem todo o Agente que o faz é punido; para que isso seja admissível ele terá, também, de ser capaz de ter Metacognição.

Adicionalmente, e porque não seria admissível falarmos de VA sem mencionar as questões éticas associadas à sua programação, dispensaremos algumas palavras neste tema, com o intuito de verificar se nos pode ajudar a determinar um regime legal para esta entidade. É também esta a finalidade da análise dos Relatórios da UE sobre este mesmo assunto.

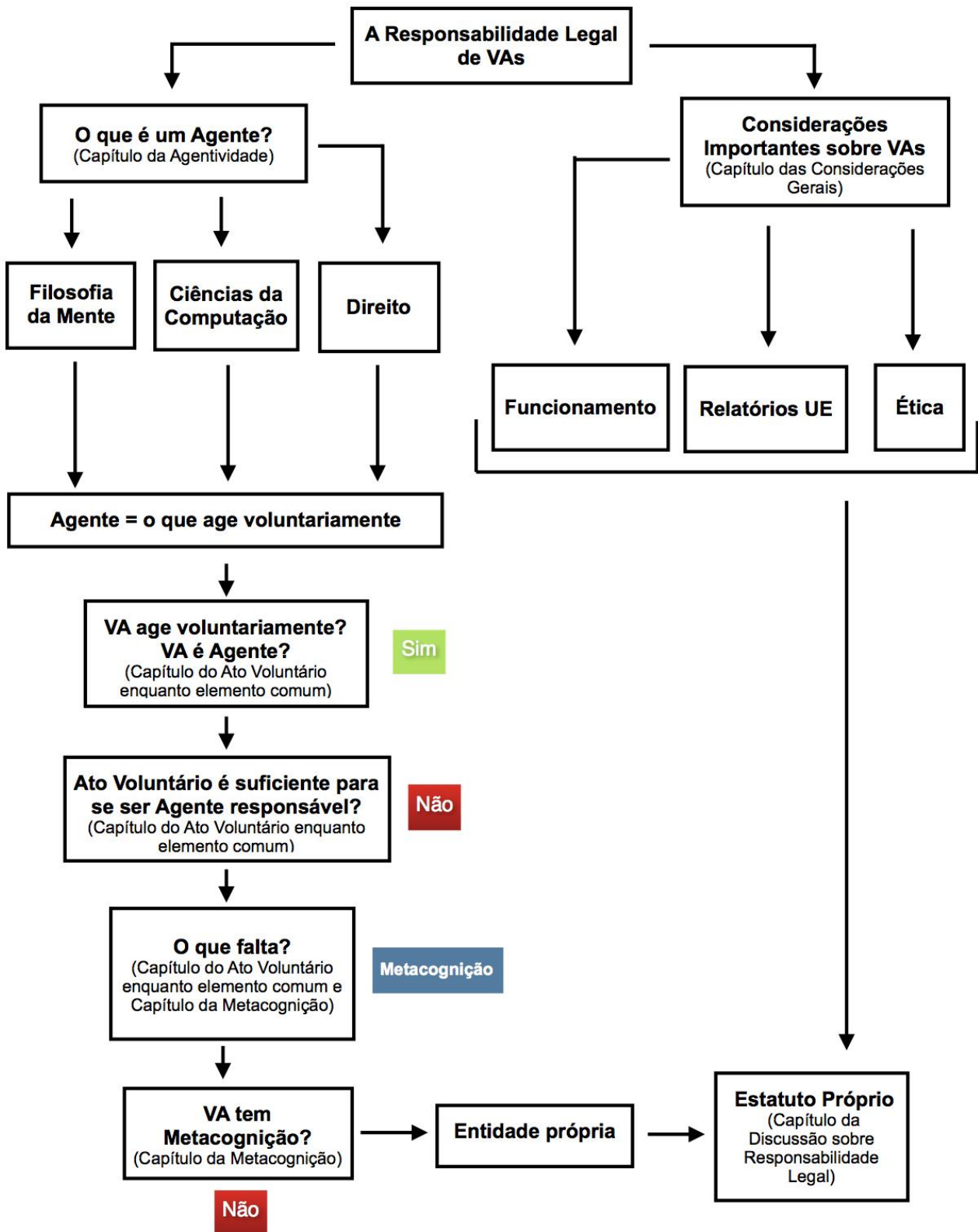
Uma nota final antes de, efetivamente, iniciarmos esta tese: para considerações feitas neste documento e tendo em conta o trabalho desenvolvido pela *SAE International*⁴, olharemos para o VA e para a sua responsabilidade em duas situações distintas: no nível 3 de Autonomia, em que o ser humano ainda pode ser chamado a intervir em caso de necessidade, e no nível 5 de Autonomia, em que o ser humano não é mais uma figura de relevo mas é meramente transportado pelo VA.

SAE level	Name	Narrative Definition	Execution of Steering and Acceleration/Deceleration	Monitoring of Driving Environment	Fallback Performance of Dynamic Driving Task	System Capability (Driving Modes)
Human driver monitors the driving environment						
0	No Automation	the full-time performance by the human driver of all aspects of the dynamic driving task, even when enhanced by warning or intervention systems	Human driver	Human driver	Human driver	n/a
1	Driver Assistance	the driving mode-specific execution by a driver assistance system of either steering or acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task	Human driver and system	Human driver	Human driver	Some driving modes
2	Partial Automation	the driving mode-specific execution by one or more driver assistance systems of both steering and acceleration/deceleration using information about the driving environment and with the expectation that the human driver perform all remaining aspects of the dynamic driving task	System	Human driver	Human driver	Some driving modes
Automated driving system ("system") monitors the driving environment						
3	Conditional Automation	the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task with the expectation that the human driver will respond appropriately to a request to intervene	System	System	Human driver	Some driving modes
4	High Automation	the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene	System	System	System	Some driving modes
5	Full Automation	the full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver	System	System	System	All driving modes

Fig.2 - Os graus de Autonomia do VA, segundo a SAE International.

Por fim, para melhor compreensão desta dissertação, deixa-se um esquema representativo da estrutura da mesma na próxima página.

⁴ Society of Automotive Engineers, que, sendo uma associação profissional Americana, determina grande parte dos padrões a seguir na indústria.



Considerações Gerais sobre os Veículos Autónomos

1. O seu funcionamento

Antes de mais desenvolvimentos, importa começar por definir o que é um VA. Procurando, não uma definição perfeita, mas apenas de uma que seja operacional, preferimos dizer que um Veículo Autónomo é um veículo que consegue conduzir sem intervenção humana (Ilková & Ilka, 2017). Se pretendermos ser um pouco mais completos nesta descrição, podemos alegar que é uma máquina capaz de atuar, ou seja, conduzir, com garantias de segurança, enquanto monitoriza as condições da estrada e do que o rodeia, durante o período de uma viagem, assumindo o controlo da mesma.

Os argumentos a favor da tecnologia são inúmeros. Com efeito, acredita-se que o trânsito deverá diminuir, a utilização de combustíveis fósseis deverá ser muito menor ou extinguir-se totalmente, as pessoas com incapacidades visuais ou motoras terão mais facilidade em movimentar-se na cidade e a tecnologia dará, ainda, a oportunidade aos indivíduos para tornarem o tempo em que estão na estrada algo com maior utilidade (Zohn, 2015).

Um dos maiores, se não, efetivamente, o maior argumento para a utilização de VAs, é o argumento da segurança (Hicks, 2018). Em abstrato, é uma ideia que faz sentido; os VAs irão ser capazes de eliminar os erros de condução, bem como travar mais rapidamente em caso de necessidade (Fagnant e Kockelman, 2013). Segundo o relatório de 2009, da *EuroSafe, Injuries in the European Union (2005-2007)*, que analisa os dados publicados pela OMS e Eurostat de 27 países de União Europeia (UE), em média, os acidentes (intencionais e não intencionais) foram responsáveis por 255.850 mortes por ano. Sem prejuízo da melhoria das condições das estradas, que podem ter um impacto significativo neste número, a tecnologia dos VA é, neste sentido, bem-vinda.

O argumento da segurança ajuda, também, na preparação do mercado e na aceitação pública deste tipo de produtos, dado que, por muito conveniente que seja um produto, as pessoas não o comprarão se não acharem que é seguro (Zohn, 2013).

O funcionamento computacional de veículos autónomos não é uniforme, visto que não se trata, ainda, de uma tecnologia totalmente estabelecida. Acresce que nem sempre os avanços nestas tecnologias são publicitados pelos produtores, devido a sensibilidades e vantagens comerciais (Katrakazas, 2015).

De uma forma global, podemos dizer que entre os métodos mais utilizados, nos vários módulos de funcionamento, se incluem as redes neuronais, sistemas de lógica difusa, algoritmos genéticos e sistemas baseados em regras (Czubenko, Kowalczyk e Ordys, 2015).

Contudo, nesta exposição, tentaremos fazer um apanhado dos seus processos de recolha de informação e tomada de decisão e apenas no indispensável para a compreensão da tecnologia.

Como tem vindo a tornar-se cada vez mais claro, as tecnologias descritas no âmbito do VA têm como principal inspiração a inteligência humana (Lonita, 2017). No sentido desta inspiração, o sistema a utilizar tem de ser capaz de agir de forma autónoma, bem como aprender quando ensinado e com a sua própria experiência e, ainda, ser capaz de fazer raciocínios aproximados (o que inclui mais do que a lógica bivalente, de verdadeiro ou falso), para lidar com contextos onde a incerteza é quase constante.

Um exemplo onde esta inspiração é notória é a classificação feita por Tadiou, em cinco categorias, a saber: (1) a resolução de problemas, (2) a representação formal (através da programação dos conceitos - Russel e Norving - por forma a que possam ser manipulados pelo VA) do conhecimento, (3) *Machine Learning* (processos de aprendizagem da máquina), (4) Processamento da Linguagem natural (porque, para interagir com o nosso mundo, os VA têm necessariamente que entender a nossa linguagem) e (5) Redes Neurais Artificiais.

O maior desafio no domínio do desenvolvimento de VAs é a análise em tempo real do ecossistema em torno do veículo e a capacidade de dar uma resposta imediata, nunca esquecendo o objetivo final que é chegar ao destino (Bayouth, Nourbakhsh e Thorpe, 1997). Por outras palavras, é preciso executar, corretamente, todas as tarefas necessárias à condução bem-sucedida, com a menor quantidade de erros possível (Lonita, 2017), visto que conduzir é uma atividade potencialmente perigosa.

Tendo analisado vários modelos existentes da arquitetura funcional de VAs e fazendo um esforço para resumir as operações que tomam lugar, Pendleton et al. (2017), dividem a arquitetura do VA em três camadas: Percepção, Planeamento e Controlo, conforme descrito na figura abaixo. As arquiteturas em camadas têm vindo a ser cada vez mais utilizadas porque permitem a partilha de informação entre os diversos módulos de processamento existentes, enquanto operam constantemente e em paralelo.

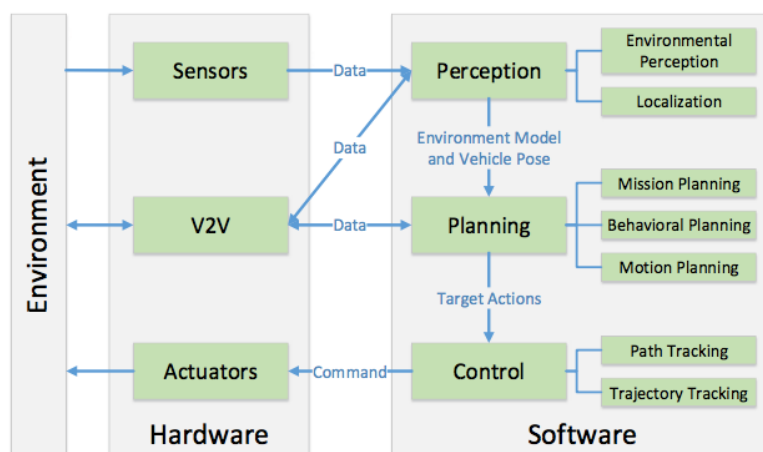


Fig. 3 - Funcionamento global de um VA, descrito por Pendleton et al. (2017).

O esquema aqui representado constitui uma versão bastante simplista do funcionamento decisório do Veículo Autônomo. Diferentes empresas terão diferentes abordagens neste âmbito, mas as ideias base encontram-se descritas nesta imagem.

1.1. A percepção

1.1.1. A percepção Interna e Ambiental

A percepção é a capacidade de o sistema autónomo selecionar e recolher informação relevante do ambiente que o rodeia (Pendleton et al., 2017). Este processo é feito através dos sensores, e é graças a eles que o sistema está apto a criar um modelo do mundo - o tipo de clima, o estado das estradas, a intensidade do trânsito. Por outro lado, é também através dos sensores, neste caso internos, por oposição aos anteriores, que são externos, que o VA conhece a sua própria posição, velocidade e a pressão dos seus pneus. Poderá ainda ter a capacidade de detetar o uso de cintos de segurança dos passageiros (até porque, hoje, quase todos os carros já o fazem), bem como o seu peso, distribuição e até o movimento dos olhos do condutor (Kalra, Anderson e Wachs, 2009). Quanto mais informação o VA obtiver, mais completo será o seu modelo do mundo.

Já a percepção ambiental refere-se ao desenvolvimento de um contexto específico e compreensão do ecossistema no qual o VA está envolto (perceber onde há obstáculos, que sinais de trânsito estão presentes, velocidades dos objetos que se encontram à sua volta e, ainda, previsão de estados futuros desses objetos), excluindo deste tipo de percepção os elementos internos (Pendleton et al., 2017).



(a) Imagens representativas da perspetiva interna do VA.



(b) Imagens representativas da perspetiva externa do VA.

Figs. 4, 5, 6 e 7 ((a) e (b)) - Imagens representativas dos VA e de como é criado o modelo do mundo (Fridman et al., 2017).

Trata-se, como se compreende, de uma função fundamental nos VAs, dado que fornece informação essencial para a condução, incluindo um elemento chave: a interpretação de sinais de trânsito (Czubenko, Kowalczyk e Ordys, 2015).

Ora, no ser humano, grande parte da informação que chega ao nosso cérebro e nos permite criar um modelo do mundo que nos rodeia é visual. Assim, no VA, importa ter um meio equivalente à visão do ser humano. Aqui teremos a utilização de câmaras, radares e lidars, sendo que, cada vez mais e considerando os resultados positivos que se têm vindo a obter, se juntam os três para um processamento de imagem mais eficaz. As fases do processamento de imagem são: a aquisição contínua de imagens, o seu processamento, a extração de traços essenciais e sua classificação, deteção e reconhecimento (Lonita, 2017). É aqui que entra o *Deep Learning*, que é um método subsumido pelo *Machine Learning*.

Carey Nachenberg (Google e UCLA, Departamento de Ciências da Computação) descreveu o *Machine Learning*, num seminário em Standford, há cerca de três anos, como uma classe de algoritmos que aprende os conceitos, automaticamente, através da análise de uma grande quantidade de dados. A noção de *Deep Learning*, por sua vez, passa, essencialmente, por dois aspetos: (1) modelos de computação com várias camadas que permitem o processamento não linear de informação e (2) independentemente da aprendizagem com ou sem supervisão, implicam o reconhecimento de traços em graus cada vez mais abstratos (Il Deng, Dong Yu, 2014).

Nos VAs, o *Deep Learning* vem colmatar a inexistência dos métodos de organização da informação que o ser humano tem, que estão ausentes na máquina. A nossa visão, transmitindo ao cérebro a informação, permite-nos identificar se o obstáculo que está à nossa frente é um peão ou apenas um saco de plástico. No caso do VA, tem que existir um método de reconhecimento dos traços característicos do objeto e uma identificação da sua categoria, para que uma decisão seja tomada em relação a esse objeto: caso seja um peão, o VA deverá parar, na identificação do objeto como sendo um saco de plástico, o VA deve continuar o seu percurso.

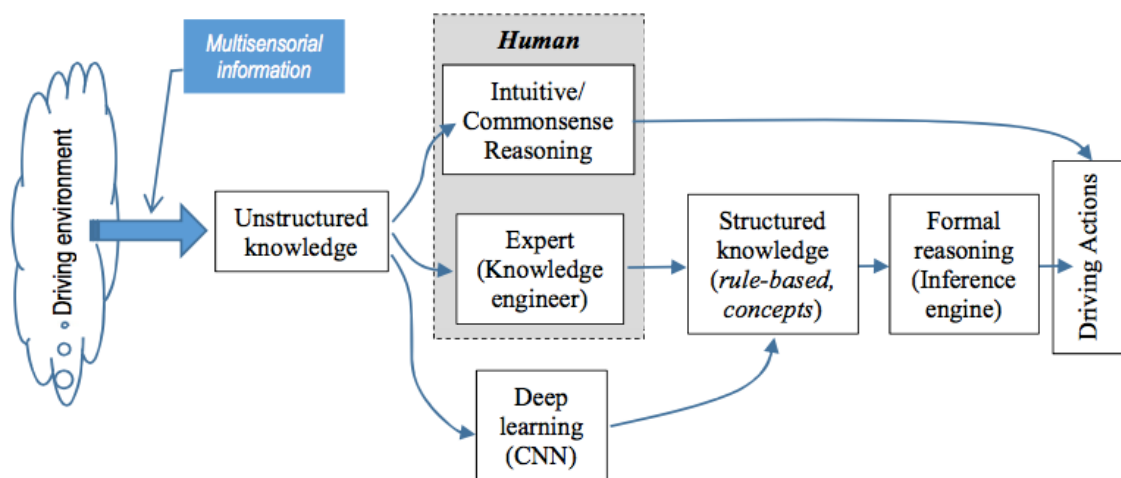


Fig. 8 - Descrição do papel do Deep Learning num quadro geral segundo Lonita (2017).

O método descrito do *Deep Learning* baseia-se nas Redes Neurais, que por sua vez encontraram inspiração na biologia e anatomia do nosso próprio sistema nervoso.

Conforme é visível no esquema da figura 9, em baixo, existem várias camadas: uma de *input* (informação dada ao algoritmo), as de *hidden layers* e uma de *output*, através do qual o algoritmo nos dá a resposta. Cada círculo representa um nódulo, cada um deles com um valor, e a ativação do próximo nódulo vai depender dos pesos (*weights*) atribuídos (Schmidhuber, 2015). Este mecanismo possui a particularidade de, caso o *output* não se demonstre o desejado, alterar os valores dos pesos, sem mão humana, até obter o resultado pretendido. É, assim, um mecanismo de aprendizagem, mas, mais que isso, de decisão autónoma. O ser humano não sabe, a certa altura, que correções é que o sistema fez, nem porque as fez.

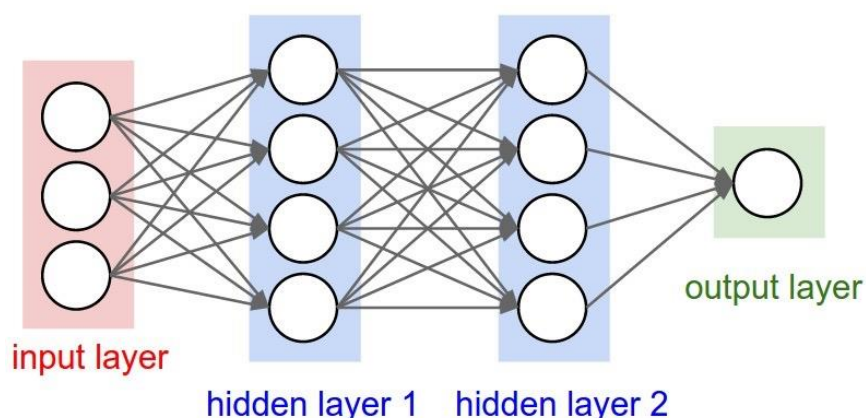


Fig. 9 - Representação esquemática de uma Rede Neuronal, acedido através de <http://cs231n.github.io/convolutional-networks/>, no dia 8 de novembro de 2017.

Sintetizando, e utilizando o exemplo de reconhecimento de imagens, o que ocorre é que o algoritmo agarra no *input*, nos traços característicos que foi treinado para reconhecer, nas probabilidades relativas que já conhece desses traços, com as quais joga nas *hidden layers*, e determina a figura global que lhe foi pedida para reconhecer.

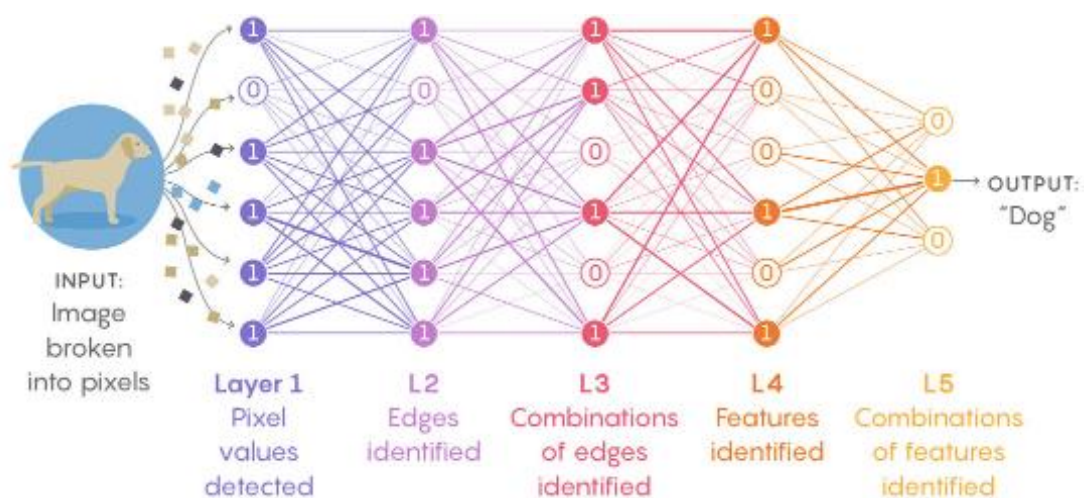


Fig. 10 - Reconhecimento de imagens através de Redes Neurais, obtido através de <https://www.quantamagazine.org/machine-learning-confronts-the-elephant-in-the-room-20180920/>, no dia 5 de outubro de 2018.

Regressando um pouco atrás, importa perceber como é que estes traços característicos, indispensáveis ao reconhecimento de imagens, são adquiridos. A primeira tecnologia que iremos descrever é o LIDAR. O LIDAR é um sistema de detecção de luz e alcance, através de milhares de pulsares de luz por segundo, medindo a reflexão para determinar a distância, sempre num eixo rotativo, fornecendo um mapa dinâmico e tridimensional do ambiente em volta do VA (Lonita, 2017). O LIDAR é, deste modo, o mecanismo mais importante no que diz respeito à detecção de obstáculos. Contudo, é um sistema que falha quando os sinais são dispersos, possuem pontos da sua caracterização ausentes ou não estão, de algum modo, organizados. Ainda assim, o LIDAR é capaz de detetar a diferença entre um saco de plástico e um pequeno animal a distâncias consideráveis (Pendleton et al., 2017).

As técnicas mais utilizadas de representação através do LIDAR são:

- a. Grelha - que divide o espaço em conjuntos mais pequenos, cuja maior dificuldade consiste em delimitar estes conjuntos do ponto de vista da eficácia;
- b. Traços essenciais - que extrai das nuvens de pontos os traços essenciais, delineando linhas (Sack e Burgard, 2014) e superfícies (Oliveira, Santos, Sappa e Dias, 2015), sendo a mais eficaz do ponto de vista do processamento, mas a mais abstrata, podendo acarretar consequências ao nível da determinação do objeto;
- c. Nuvem de pontos - que usa diretamente a informação do sensor para processamento, gastando mais memória e tempo de processamento (Asvadi, Premebida, Peixoto e Nunes, 2016).

Em todo o caso, o processamento através do LIDAR implica sempre mecanismos de segmentação e classificação da imagem recolhida.

Além do LIDAR, outros sistemas equivalentes à visão humana são utilizados, como câmaras, nomeadamente, para detetar as linhas das faixas rodoviárias, o espaço livre de circulação (sem que haja colisão), os veículos e os peões no ambiente envolvente. A fusão dos dois sistemas tem vindo a ser fortemente encorajada.

A literatura sobre VA não dispensa, ainda, utilização de Radares, que emitem ondas rádio sucessivas para detetar objetos num raio considerável, bem como as suas velocidades.

Para terminar, é de referir que na IA, em termos globais, são muitos os avanços que têm vindo a ser feitos neste âmbito da visão. Em Agosto de 2018 foi apresentado um sistema capaz de detetar em TACs, com 95% de precisão, a existência de tumores, isto face aos 65% de precisão, que os seres humanos são capazes de ter (<https://www.sciencedaily.com/releases/2018/08/180822114436.htm>).

1.1.2. A Localização

Este aspeto refere-se à capacidade de o veículo determinar onde se encontra em relação ao ambiente, bem como de medir o seu próprio movimento. A técnica, aos dias de hoje, com maior popularidade tem sido a utilização do GPS e, adicionalmente, porque a margem de erro do GPS é considerável, sistemas de navegação

por inércia. Estes utilizam instrumentos como os acelerómetros, giroscópios e técnicas de processamento de sinais, que não necessitam de aparelhos externos ao veículo, calculando o movimento, orientação, inclinação e altitude do veículo.

Mais concretamente, no que diz respeito aos sistemas de navegação por inércia, temos métodos como (Pendleton et al., 2017):

- a. Odometria - que, sabendo a posição inicial do veículo, o tempo da viagem e o deslocamento das rodas, consegue calcular, com precisão, a localização do veículo.
- b. Filtragem de Kalman - utilizado em sistemas dinâmicos (que evoluem no tempo, como é o caso do VA), tem a capacidade de agregar a informação recebida de vários sensores sobre o mesmo objeto; neste método, existe sempre uma previsão retirada do modelo do mundo conhecido, e alternância entre esta previsão e correção/atualização da informação.
- c. Filtragem de Partículas - que compara, com o mesmo objetivo do mecanismo anterior, as observações dos vários sensores com o mapa, criando partículas à volta das zonas do mapa que se sobrepõem à informação recebida pelos sensores.
- d. SLAM (*Simultaneous Localization and Mapping*) - dos métodos mais utilizados nos VA e na robótica em geral, que estima a posição atual do veículo bem como de objetos de referência (tais como semáforos, casas, edifícios).

1.2. O Planeamento

No domínio dos VAs, entende-se planeamento como o processo de tomada de decisões, decisões estas que, naturalmente, não são aleatórias, mas que têm um propósito, sendo um meio para atingir um fim. Em termos específicos quer isto dizer que é aplicado um algoritmo para analisar o modelo do ambiente, determinando as ações a executar (Kalra, Anderson, Wachs, 2009) em harmonia com esse modelo desenhado previamente pelo VA. Aqui irão incluir-se, idealmente, nesta conceção do mundo, todos os comportamentos que um condutor normal pratique tais como mudar de faixa para em seguida virar à direita, virar mais ou menos o volante, tudo isto a uma determinada velocidade, etc. O planeamento é uma fase que consome muita memória, que é computacionalmente intensa e que acontece em paralelo com outras operações do veículo, como fusão de informação, deteção de obstáculos e módulos de controlo (Katrakazas, 2016).

Dentro da ideia geral de planeamento, existem processos bem específicos, entre os quais se enumeram alguns:

- a. Planeamento de Missão - no qual o VA decide por que estrada ou estradas optar para chegar ao destino. O sistema representa as redes de estradas como um gráfico dirigido, com os valores associados aos custos de enveredar por cada caminho disponível, sendo o objetivo a determinação da solução ótima (Paden et al., 2016), isto é, a solução mais eficaz tendo em conta os custos de cada opção. Na prática, o que ocorre é que diferentes estradas terão diferentes valores, que podem ser calculados, por exemplo,

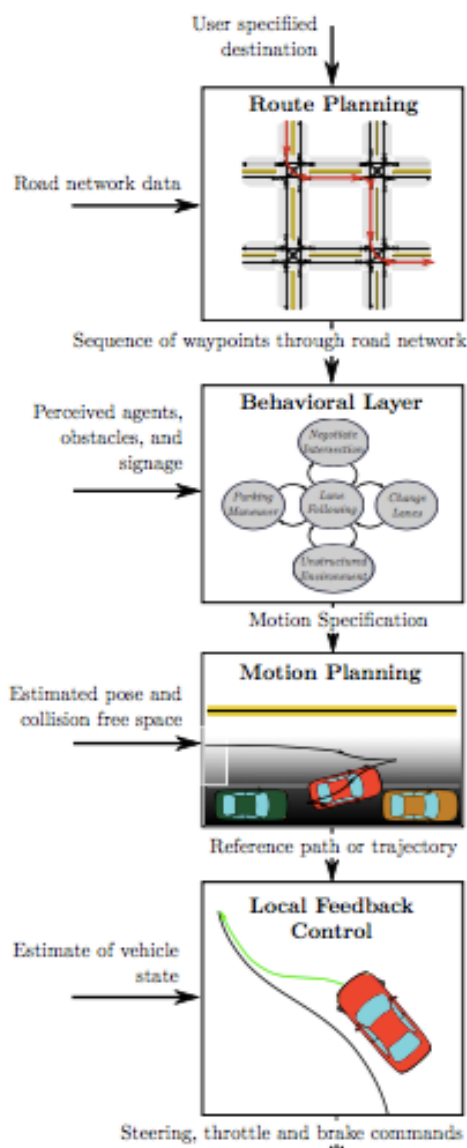


Fig. 11 - Descrição hierárquica do planejamento (Paden, 2016).

controle de *feedback* para escolher o efetor apropriado para levar a cabo a decisão da fase de planejamento de movimento e corrige erros que possam surgir, durante o movimento, devido a alguma imprecisão do modelo de veículo. Em concreto, o que esta fase faz é, nomeadamente, ordenar ao volante a direção, a pressão ideal aos pneus e os piscas a utilizar.

1.3. O controle

Trata-se da capacidade do Robô de executar as ações planejadas, passando das intenções e decisões à ação propriamente dita, corrigindo, por exemplo, a trajetória, comparando a posição do carro e a posição desejada. Importa, também, que o sistema consiga rejeitar uma leitura errada que um outro módulo tenha feito, como seria o caso de um módulo identificar uma perturbação no ambiente que, na realidade, não existe. Estas rejeições, por sua vez, alteram toda a dinâmica do sistema, que havia sido previamente determinada, garantindo, contudo, que se atinge o estado desejado, que é a condução em segurança até ao destino. O mecanismo mais

em função dos quilómetros a percorrer, pela intensidade do trânsito ou possibilidade de existir um acidente.

b. Planeamento comportamental - após a determinação de um caminho, o VA tem que estar apto a cumprir as normas do código da estrada e ser capaz de ter uma interação adequada com outros Agentes de forma segura, ao mesmo tempo que há uma progressão do ponto de vista da missão. Por exemplo, se o VA está a chegar a uma interseção, é esta camada que implica a sua paragem, observa os veículos e peões circundantes e determina que é a sua vez de passar (Paden et al., 2016). Tendo em conta a incerteza inerente ao comportamento dos restantes veículos na estrada, são utilizadas técnicas de *Machine Learning* e planeamentos probabilísticos para dar conta dessa incerteza.

c. Planeamento de movimento - tendo em conta a decisão na camada anterior, o planeamento de movimento afina a decisão, por forma a garantir o conforto do passageiro e a não colisão com obstáculos presentes. Por outras palavras, é a parte responsável por computar uma trajetória segura, confortável e dinamicamente possível, dado a configuração de objetivos feita pela camada comportamental (Paden et al., 2016). Por exemplo, se o comportamento definido for estacionar num lugar, esta fase vai pegar na informação que tem sobre os obstáculos que envolvem o VA, quer dinâmicos quer estáticos, e gera a trajetória ideal, tendo sempre em conta as restrições dinâmicas e cinemáticas do veículo.

d. Planeamento dinâmico ou de veículo - esta camada usa um

utilizado é o controlo de *feedback*, que responde valorativamente e compensa desvios para atingir o comportamento desejado.

Para completar este sistema, e para otimização de resultados, são utilizados, com frequência, modelos preditivos (cuja ideia base se encontra enunciada no esquema) que permite a existência de um controlo de *feedback* de diversas variáveis.

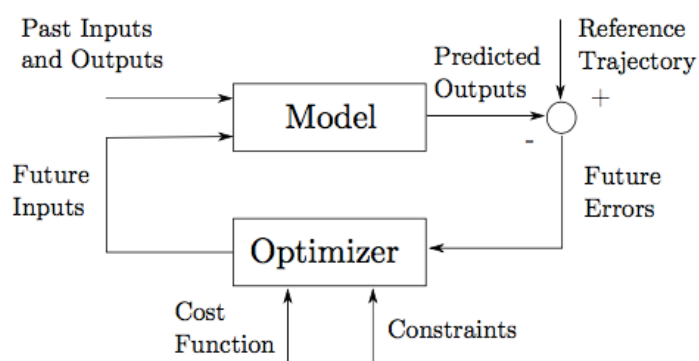


Fig. 12 - Estrutura básica do Modelo Preditivo de Controlo de acordo com Pendleton et al. (2017).

O que este tipo de controlo faz, em poucas palavras, é resolver o problema de otimização da trajetória, no tempo limitado que tem, de maneira recursiva, tendo em conta a atualização da informação que está constantemente a acontecer.

O âmbito do controlo é, portanto, a geração de trajetórias, com a informação anteriormente recolhida, assim como a verificação de que esse caminho está a ser devidamente seguido. O sistema de controlo é capaz de prever estados futuros do ambiente, utilizando, também, estes dados, para a otimização. Deste modo, no quadro geral, este módulo é também essencial para evitar qualquer colisão.

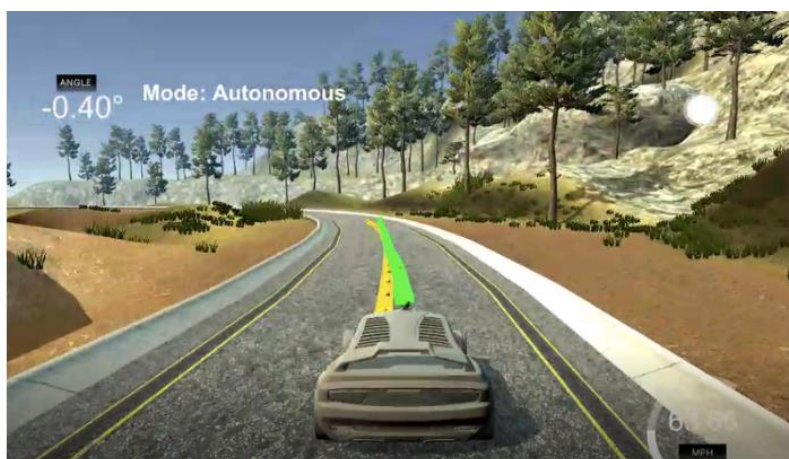


Fig. 13 - Comparação de Trajetórias (a amarelo vs a verde). Imagem obtida através de <https://medium.com/@techreigns/model-predictive-control-implementation-for-autonomous-vehicles-932c81598b49>, acedida no dia 7 de Outubro de 2018.

É de referir que as técnicas de *Deep Learning* têm, hoje, uma capacidade de predição já bastante avançada. O ser humano consegue prever pequenas situações concretas, ainda que racionalmente não as consiga explicar e isto permite evitar bastantes acidentes. Também o VA é capaz de prever a trajetória dos restantes veículos, fazendo-o através da informação que recolhe com os seus sensores sobre o seu estado e o dos restantes objetos (Singh, 2018).

Para esta previsão, além do modelo que referimos acima, no qual o VA, por aprendizagem, vai criando e melhorando o seu próprio modelo, atribuindo probabilidades aos diversos comportamentos possíveis dos outros condutores, há ainda uma abordagem que se serve de uma quantidade enorme de dados sobre os comportamentos dos condutores, em treino *offline*, antes de aprender com dados reais (Singh, 2018). Assim, na vida real o VA observa a trajetória parcial do outro veículo, compara-a com as trajetórias dos protótipos através dos quais aprendeu e é capaz de criar uma previsão.

Para melhores resultados de previsão o processo mais utilizado é um híbrido entre estas duas abordagens, utilizando a criação de modelos do primeiro e juntando com o *Machine Learning* do segundo, trazendo capacidades de previsão bastante precisas.

Antes de terminarmos este capítulo, há um mecanismo importante a referir, que entra no subcapítulo do controlo e predição. Uma das formas mais eficazes de controlo dos objetos circundantes e que permite esta previsão de estados futuros dos diversos elementos na estrada é a chamada comunicação V2V (*Vehicle-to-Vehicle* - Veículo para Veículo). Esta tecnologia permite a partilha da posição atual, e previsão da posição futura entre todos os veículos existentes num dado momento na via. Estas comunicações são, com certeza, uma mais valia, na tarefa de evitar a colisão com outros carros na via (Levy, 2016, através do *Insurance Institute*), apesar de todas as questões legais de proteções de dados que devem ainda ser resolvidas.

Já havia sido referido que os métodos e mecanismos utilizados variam de empresa para empresa, pelas mais diversas razões. Por outro lado, descrever detalhes técnicos implicaria tornar a dissertação bem mais extensa, sem que benefícios pudessem daí ser extraídos, dado que pretendemos apenas o mínimo que nos permita concluir sobre a capacidade dos VA de agirem voluntariamente.

Terminada esta descrição de VAs que, conforme já mencionado, é assumidamente simplista, não pretendendo ser mais do que isso, iremos avançar para outras considerações gerais sobre VAs. Resta-nos olhar para o relatório da UE e fazer algumas considerações Éticas sobre os mesmos.

2. Análise do Relatório da União Europeia sobre Veículos Autónomos

A União Europeia preparou um documento, *A Common EU Approach to Liability Rules and Insurance for Connected and Autonomous Vehicles*, no qual são feitas várias reflexões e orientações sobre o regime jurídico do VA e do qual nos serviremos como ponto de partida para determinação de um regime.

Logo de início, e baseando esta visão em relatórios finais de estudos, tais como *GEAR 2030 HLG*, datado de dez de outubro de 2017, a UE afirma ser consensual que legislação em vigor, nomeadamente no que diz respeito à utilização de seguros e à diretiva de responsabilidade por produto defeituoso, se adequa ao contexto em que os VAs se inserem. Até admitimos que se compartilhe, em larga medida, este entendimento, mas é prudente fazer uma análise de antemão, dado o rápido desenvolvimento destas tecnologias. Até uma alteração legislativa neste âmbito, estaremos reféns de uma apreciação casuística do regime, o que não é, como já vimos, desejável.

A título de exemplo, podemos concordar que uma Diretiva Relativa à Responsabilidade Decorrente dos Produtos (doravante DRRDP) seja suficiente para assegurar uma distribuição equilibrada da responsabilidade no caso de defeito num VA. Cabendo, o VA, na categoria de produto colocado num mercado por uma dada empresa, a mesma será responsável pelos defeitos que este venha a apresentar, num prazo considerável. Contudo, inúmeras questões se acrescentam à equação quando falamos de responsabilidade pela utilização. No entanto, e ainda respeitante à DRRDP, o que é interessante notar são as novas vertentes de responsabilidade que os VAs vêm acrescentar aos seus produtores. Logo, esta diretiva precisa, indubitavelmente, de reflexão e estudo sobre a sua aplicabilidade.

Quais serão, então, as questões mais prementes e sobre as quais urge uma resposta? A União Europeia destaca três pontos absolutamente fulcrais, a um nível abstrato, a serem analisados, quando o tema é Responsabilidade Civil em Veículos Autónomos:

- (1) As limitações da responsabilidade - isto é, o alcance da recuperação possível de danos e formas de compensação;
- (2) O sistema de determinação da responsabilidade - ou seja, se deve ser adotado um sistema de responsabilidade objetiva ou de responsabilidade pelo risco, com limites delineados em proporção do nível de autonomia dado ao VA, bem como às instruções dadas;
- (3) A obrigatoriedade de seguro e fundo de garantia afeto à utilização do Veículo.

Por outro lado, e num plano mais concreto, a UE antecipa os seguintes problemas:

- (1) O sentido restrito de defeito, que pode ser difícil de alargar para os VA, em toda a sua complexidade e técnica. Segundo o estudo, todos os danos decorrentes da utilização na estrada, má reparação e condições atmosféricas não estão cobertos pela DRRDP. Sabendo que os produtores e distribuidores tudo farão para se ilibar de qualquer tipo de responsabilidade, esta situação deixa o consumidor imensamente desprotegido. Lançada está, também, a questão sobre se um *software* é um produto; embora a UE ainda não tenha determinado uma definição legal de produto, noutra tipo de tecnologias, como nos computadores e dispositivos médicos, o *software* sempre foi visto como um produto, pelo que não haverá razão para considerar outra coisa.
- (2) O ónus da prova dos riscos cientificamente não conhecidos, no regime tal qual como hoje se apresenta, recaem sobre a parte que sofre o dano. Gera-se, assim, a questão de sabermos como se podem provar danos resultantes de causas que não se conhecem, dada a complexidade da tecnologia de que falamos.

Podemos colocar a hipótese do meio de prova através de peritagem, mas continuamos sem saber o nível de conhecimento que tem que ter um perito para ser considerado como tal, principalmente tendo em conta as diferentes arquiteturas existentes nas variadas empresas do mercado. Por outro lado, obrigar as seguradoras ou empresas de produção a ter um perito disponível para prestar declarações em tribunal não é o método mais transparente e imparcial. Adicionalmente, teríamos que considerar os custos de se trazer um perito a tribunal, que recaem, ainda assim, no lesado. Para este efeito, seria de considerar inverter o ónus da prova, visto que é mais fácil o produtor provar que não houve falha sua.

Das maiores dificuldades ao momento é o facto de a lei vigente em cada país divergir dos restantes. Não obstante, estão bem representados os regimes baseados na responsabilidade subjetiva (na culpa do condutor), responsabilidade objetiva, responsabilidade pelo risco e, ainda, regimes mistos. O tronco comum em todos estes sistemas de determinação da responsabilidade, contudo, é a assunção de que o condutor está em controlo do veículo. Com os VAs, esta assunção basilar pode mudar. E terá, necessariamente, que mudar.

Mais se nota que os regimes em que a responsabilidade se relaciona com a culpa do indivíduo, podem, com esta tecnologia, vir a ser considerados totalmente obsoletos, dados que a decisão poderá não vir a ser imputada ao utilizador do veículo que, com o passar dos anos, deixou de ser o condutor, e é visto como mera carga.

Tendo sido colocadas cinco questões pertinentes sobre o regime de responsabilidade, três abstratas e duas concretas, às quais cumpre responder, o relatório da UE passa a enumerar os riscos que dão origem a estas questões. Por outras palavras, tratam-se das fontes da responsabilidade, a saber:

- (1) Falhas do sistema operativo do VA - aqui importa perceber quando e sob que condições o fabricante do *software* se deve responsabilizar, e, logo, quando é que se exclui o fabricante da própria mecânica do carro. Ou seja, é relevante determinar de forma clara o que é *software* do VA e o que é a sua mecânica. Adicionalmente e como vimos, é imperativo perceber em que circunstâncias é que a falha do *software* pode ser considerado sob a alçada da DRRDP, se é que pode, de todo. Nesse caso, então, descrever o que significa um *software* defeituoso e contra que produtor, o do *software* ou do carro, deve o utilizador do VA propor a ação de responsabilidade. De acordo com a UE, tal irá depender das razões para o sistema ter falhado. Este tipo de riscos está assegurado na medida em que estas falhas poderiam ou deveriam ser cientificamente conhecidas antes do VA ser comercializado, sem prejuízo de legislação nacional sobre o tema.
- (2) Falhas na rede - este item pressupõe a dependência do VA de uma rede de internet e considera os problemas dessa mesma rede. Aqui importa saber: o fornecedor da rede pode vir a ser considerado responsável? A resposta a esta pergunta depende do que se considera como VA propriamente dito e o que é acessório externo. Sendo considerado parte do VA, o produtor deste será, naturalmente, responsável direto; em caso negativo, deve determinar-se o responsável.
- (3) Riscos relativos ao *hacking* e crime cibernético - parece claro, como também indica o relatório da UE, que *hacking* e acesso indevido não estão abrangidos pela legislação relativa aos carros com condutor. Caso haja um defeito em termos de proteção, anterior à colocação no mercado, o produtor pode vir a ser considerado responsável, uma vez que esta tecnologia deve ter mecanismos de defesa contra este tipo de

perigos. Uma das dificuldades poderá estar, também aqui, como no ponto um, na prova deste defeito. A outra será na atribuição de responsabilidade quando todas as precauções foram tomadas e, ainda assim, se deu o crime. Pode ser útil definir-se padrões mínimos de segurança que têm que ser cumpridos. Contudo, a acontecer, estes padrões devem ser atualizados periodicamente, visto que a tecnologia, para o bem e para o mal, evolui a uma velocidade considerável.

- (4) Riscos relacionados com as escolhas de programação (incluindo as escolhas éticas) - Finalmente, no caso das escolhas feitas pelo programador, a questão é saber se poderá a programação ineficaz ou incorreta ser considerada defeito de *conceção* e, logo, um produto defeituoso. Esta é mais uma questão que a DRRDP não consegue responder.

A opinião, ao momento, da UE, é que a hipótese mais viável seria criar um instrumento totalmente novo de regulação, para abranger os novos riscos, incluindo o seguro de responsabilidade objetiva.

2.1. Relatório Final apresentado à União Europeia por Dr. E.F.D. (Esther) Engelhard and R.W. (Roeland) de Bruin, LL.M. (anexo ao *A Common EU Approach to Liability Rules and Insurance for Connected and Autonomous Vehicles*)

Este relatório, além de procurar responder às questões anteriores, visa apresentar as vantagens e desvantagens de cada regime, permitindo-nos refletir sobre qual será o mais adequado tendo em conta a tecnologia que temos em mãos.

É sabido que a DRRDP impõe um regime de responsabilidade objetiva pelo defeito de determinado produto. Do nosso ponto de vista, existindo este instrumento legal que, com as devidas alterações para incluir toda a tecnologia dos VA, pode, sem grande esforço, ser diretamente aplicável a todos os Estados Membros (EM), esta é, à partida, uma solução candidata a servir de base ao regime. Vamos começar, assim, por ver o que diz este relatório sobre a DRRDP.

Deste modo, nos termos descritos por este documento, a DRRDP seria aplicável às diversas partes que participaram na criação e publicidade do produto, bem como a qualquer entidade que se apresente como produtor (através do nome, *trademark* ou traço distintivo). Aqui podem incluir-se programadores do *software* com a sua própria marca, e abarca certamente os fabricantes dos diversos componentes do VA (rodas, sensores, *airbags*, etc.), não deixando de fora os fornecedores associados à parte tecnológica do veículo. Em suma, podemos vir a incluir qualquer fabricante, seja ele autor do *hardware* ou do *software*, exceto, conforme alerta o documento, (1) produtores de componentes relativos ao design ou à informação sobre defeitos do veículo, (2) distribuidores ou vendedores que se sirvam da tecnologia de informação *Track-and-Trace*, (3) programadores que não utilizem marcas e (4) empresas de aluguer de veículos ou prestadores de serviços como táxis. Em suma, qualquer fabricante, seja de que tipo for, estará sujeito à DRRDP.

Este facto não impede que o fabricante do *software*, enquanto componente específico do VA, tenha várias defesas à sua disposição. Produtores de componentes integrantes de um produto podem ilibar-se da responsabilidade provando que se trata de um defeito de *design/conceção* ou de informação, assim transpondo a

responsabilidade para a respetiva parte faltosa, neste caso a quem cumpria desenhar o produto desde a base ou a quem competia dar toda a informação ao consumidor.

O artigo 2º da DRRDP deixa margem para que o *software* possa, ele próprio, constituir um produto, separado do veículo como um todo. Dado que a UE tem apostado numa definição ampla de produção, este pode ser um caminho viável, embora assim seja apenas na circunstância de se conseguir vir a provar o defeito no *software*, algo que implica, como já vimos, um conhecimento técnico não ao alcance de todos, até mesmo por razões de confidencialidade. Adicionalmente, conta-se que o consumidor tenha que ser capaz de provar que o defeito já existia quando chegou a si.

Até aqui, já ficámos a saber que tipo de fabricantes podem estar sujeitos à DRRDP e já percebemos que *software* pode, e deve, ser considerado um produto. Mas o que significa dizer que o *software*, enquanto produto, estava defeituoso?

Quanto a esta definição, de produto defeituoso, o relatório sugere que um produto é defeituoso, no que se aplica aos VA, na medida em que não oferece a devida e expectável segurança, tendo em conta o caso concreto. Neste sentido, e inspirado no que já tem vindo a ser feito em relação a outros produtos, o Relatório divide categorias de defeitos, nomeadamente:

- (1) Defeito de Fabrico – defeitos que estão no veículo ou nos seus componentes e apresentam uma anomalia, impedindo de prestar funções na perfeita condição (sensores que não funcionam bem ou falha dos travões); uma definição para este defeito, na Lei Americana sobre os produtos defeituosos é a ideia de que o produto diverge do *design*/conceção planeado/a para o mesmo, apesar de todo o cuidado na preparação e *marketing* do produto. Nesta situação, e no determinado pelos precedentes do país, aqui incluem-se a utilização de partes do veículo defeituosas e a conjugação errada dos vários componentes do veículo (Kalra, Anderson e Wachs, 2009).
- (2) Defeito de Informação – os produtores têm a obrigação de fornecer a informação necessária para um uso seguro do produto; significa isto que o utilizador deve dispor de um manual de instruções e avisos quanto aos possíveis riscos e perigos; importante ressalva no facto de os produtores terem que ter em consideração que os utilizadores podem esquecer-se das instruções ou subestimar os riscos, pelo que podem ser necessários avisos adicionais. Nos termos descritos por Kalra, Anderson e Wachs (2009), caímos neste defeito quando o mesmo decorre de instruções inadequadas e quando o risco de utilização poderia ter sido minimizado ou evitado se as instruções tivessem sido bem dadas. No caso da IA, deve ser defendido um dever de informação mais forte, constando do manual de instruções os potenciais perigos daquela tecnologia e, possivelmente, as punições e/ou indemnizações aplicáveis. Assim, por exemplo, no caso dos *Drones*, cuja utilização descuidada tem vindo a colocar em causa a segurança dos aviões (veja-se o caso do aeroporto da Portela e de Gatwick) e, conseqüentemente, das pessoas a bordo deles, os perigos concretos de sobrevoar certas zonas devem estar indicados.
- (3) Defeito de *Design*/conceção – que envolvem, geralmente, um grande número ou uma série inteira de produtos, neste caso, veículos (ex. VA que foram testados em poucos cenários, não estando preparados para lidar com outros contextos fora dos previstos). Nos EUA são utilizados dois testes para verificação

deste defeito - as expectativas dos consumidores (embora alguns estados tenham abandonado este padrão, visto que muitos consumidores apresentam expectativas não razoáveis) e o custo-benefício. No custo-benefício há uma ponderação de valores em termos de perigo e probabilidade de causar dano vs benefício e facilidade de substituição do produto por parte do fabricante (Kalra, Anderson e Wachs, 2009).

Desta classificação se conclui que cada parte ou componente do VA deste estar associado a uma empresa de produção que poderá ser responsabilizada. O que será importante, no caso concreto, será, então, perceber onde está a falha.

Dois apontamentos neste âmbito importam ser feitos. A falha na segurança só existe na medida em que o produtor tenha prometido esse funcionamento a 100%, é o primeiro deles. Por outro lado, o que significa uma expectativa razoável de segurança? Será o arquétipo da segurança o condutor humano? O desenvolvimento e determinação dos termos de um teste apropriado para deteção de defeitos de *design* é uma questão central neste âmbito e não parece estar perto de ser resolvida (Vladeck, 2014). Contudo, parece lógico que temos que exigir, para a comercialização, padrões muito altos de segurança. Muito embora não seja admissível esperarmos um produto perfeito, até porque dada a complexidade da tecnologia, é uma tarefa complicada, terá que ser exigível algum tipo de segurança-padrão; é aqui que deve entrar a legislação para determinar requisitos mínimos, uma vez que a DRRDP não resolve boa parte dos novos contextos que surgem com a utilização de VAs.

O facto de ser o consumidor a provar que o produto estava defeituoso (ainda que se admita que essa prova não tenha que ser exata) em caso de anomalia, e dada a dificuldade e custos que isso pode acarretar, pode pesar de tal modo no utilizador que o irá constranger no momento de comprar um VA. Até porque, refira-se, a definição de defeito varia de Estado Membro para Estado Membro.

Para resolver os temas até agora enunciados, bem como sintetizar algumas ideias que foram defendidas, são apresentados, no relatório, sugestões de modelos de reforma, a saber:

a. Reforma da Diretiva da responsabilidade pelo produto:

a.1. Dadas as seguintes dúvidas que surgem com a aplicação da DRRDP:

- (1) *Software* pode ser considerado produto?
- (2) Sendo um produto, quando é que este se pode considerar defeituoso?
- (3) Quais os limites da responsabilidade do fabricante?
- (4) Até que ponto os novos riscos tecnológicos podem ser atribuídos à indústria (produtores não são responsáveis por riscos que não se poderiam saber, objetivamente, ou que eram de algum modo indetetáveis – *bugs*, riscos de *design* - nem por danos subsequentes)?

a.2. E os seguintes problemas:

- (1) O facto de estarmos limitados aos defeitos ou riscos que já existiam antes, excluindo os supervenientes;
- (2) Mais, o facto de estarmos, apenas, restringidos aos defeitos que são cientificamente conhecidos ou que deveriam ter sido reconhecidos antes do momento de comercialização;
- (3) Os custos técnicos e financeiros de análise dos riscos ou defeitos;
- (4) A dificuldade em determinar, normativamente, o que se inclui dentro do conceito de defeito e do conceito de culpa.

Foi proposta a amplificação do conceito de produto para incluir *software*, a substituição da noção de defeito por um critério mais restrito e a determinação de um montante indemnizatório máximo. Esta sugestão de regime, feita através de uma reforma de DRRDP, aloca a responsabilidade no fabricante.

a.3. Vantagens da alocação do risco no fabricante:

- (1) Por um lado, os produtores são quem tem o melhor acesso à informação sobre o produto e a sua segurança, são que têm mais conhecimento técnico, estando mais aptos para encontrar soluções para os problemas (Rapaczynski, 2016), logo faria sentido carregarem o risco.
- (2) Obriga a que analisem bem e confrontem os riscos/benefícios e assegurem mínimos de segurança, antes de colocarem os carros em circulação.
- (3) É mais equitativo – são os produtores que beneficiam pela colocação do produto no mercado (estes riscos podem ser considerados *business risks*).
- (4) Ao atribuir a responsabilidade ao produtor, ele pode transferir, de certo modo, o custo dessa responsabilidade, para o consumidor, através do preço de venda – aumentando-o, ainda que com as respectivas consequências (tais como menos consumidores estão dispostos a comprar, mas em compensação com menos possibilidade de acidentes).

a.4. As Desvantagens:

- (1) Passar a responsabilidade para o produtor pode retardar o desenvolvimento.

Apesar de todas estas questões que surgem, o relatório sugere que há margem para alterar e melhorar esta Diretiva, por forma a atingir a proteção e distribuição dos riscos adequada, mesmo no caso dos riscos não conhecidos. Ainda assim, no que diz respeito à colisão entre veículos, bem como aos defeitos supervenientes da viatura, à partida a solução não será encontrada na DRRDP, isto é, há que considerar estes dois regimes em paralelo.

b. Reforma das regras de trânsito (no sentido da sua harmonização):

Teria que ser feita através de *hard law*, que se tratam de diretivas vinculativas, obrigatórias para todos os EM e não de *soft law*, que transmitem, apenas, ideias globais comuns, uma vez que se tratam de ideias muito complexas cuja regulação tem que ser precisa.

Considerando que nem todos os países possuem o mesmo regime quanto à responsabilidade pelos acidentes de viação, a sugestão, neste domínio, passa por se uniformizar, optando por uma responsabilidade pelo risco, significando isto que a responsabilidade seria atribuída ao proprietário ou utilizador do veículo, independente de culpa deste. Em todo o caso, pode ser útil trazer a defesa da causa externa para a equação.

b.1. Vantagens da uniformização no sentido de responsabilidade pelo risco:

- (1) Confronta o proprietário e o utilizador com o risco associado (pessoalmente ou através de seguradora);
- (2) Oferece segurança de forma igual em todos os membros (visto que as proteções não são as mesmas quando comparamos regimes nos países em que a responsabilidade é baseada no risco, por comparação aos que se baseiam na culpa);
- (3) Protege a inovação e encoraja os produtores/fabricantes a levar a tecnologia adiante (Duffy and Hopkins, 2014, p. 119), o que confere maior competitividade à União Europeia.

b.2. Desvantagens:

- (1) Assumir este tipo de responsabilidade é assumir que o proprietário ou utilizador são responsáveis por riscos que não controlam (tal como nos veículos a motor, mas com maior complexidade por razões de autonomia do veículo), ainda que motive o proprietário e/ou condutor a ser mais cuidadoso com a manutenção, visto que tem consciência que o risco é seu.
- (2) Complexidade da harmonização. Havendo uma enorme dispersão de leis e matérias e países com regimes baseados na culpa, este tipo de reformas implicariam processos demasiado complexos de alteração, que o relatório descreve, inclusivamente, como algo inconcebível.

c. Seguro independente de culpa:

Independentemente do facto de o segurado ser o produtor ou o consumidor, o risco está diretamente associado ao seguro sem necessidade de culpa. É objeto de seguro o mero facto de o veículo ter sofrido um acidente, não ultrapassando, a compensação, os danos do acidente.

Este regime (*no-fault Insurance*) resolve o problema da responsabilidade civil através de um seguro relativo ao risco, que por sua vez está associado a um tomador do seguro. Este será, em princípio, o proprietário do veículo ou o fabricante do mesmo. Esta política irá cobrir os danos sofridos pelas pessoas no interior do veículo e possivelmente os restantes envolvidos no acidente, mas estaria, ainda assim, limitado aos riscos específicos

gerados por esta nova tecnologia inerente aos VA, sem prejuízo de se vir a criar um seguro com um âmbito mais alargado.

O relatório refere vários graus de complementaridade ou substituição do regime de responsabilidade civil com este sistema de seguros mas, a acontecer, e para ser efetivo, teria que ser obrigatório para todos os Estados Membros. Por outras palavras, pode muito bem ser útil uma conjugação entre o regime de responsabilidade civil e este seguro obrigatório.

Na aplicabilidade deste regime, importa, então, ver que papel e que termos sobram para a responsabilidade civil propriamente dita. Uma das sugestões seria a utilização do regime de responsabilidade civil na medida em que os danos excedem o seguro existente, independente de existir alteração da legislação nacional ou não. O documento em análise é insistente nesta ideia de se manter o regime de responsabilidade civil a par com o seguro obrigatório.

O relatório menciona também o caso sueco, em que os dois regimes coexistem, onde as pessoas optam, quase sempre, pelo ressarcimento de danos através do seguro, por requerer menos esforços probatórios.

c.1. Vantagens do Seguro de responsabilidade independente de culpa:

- (1) Este modelo será mais eficaz nos termos em que se espera que reduza os casos em que se recorre ao tribunal para obter compensação, dado que o lesado obtém essa mesma compensação através da sua seguradora.
- (2) Por outro lado, optar por este regime, sem prejudicar o regime de responsabilidade civil, é vantajoso porque, em conjunto, estes dois sistemas conseguem abarcar um maior número de situações concretas, o que se demonstra essencial quando as consequências de uma nova tecnologia não se conseguem prever totalmente. Um sistema dualista de compensação permitiria que os tomadores de seguro, e também as próprias seguradoras, fossem gradualmente, à medida das necessidades e do avanço da tecnologia, optando por um ou outro regime.

c.2. Desvantagens do Modelo:

- (1) A primeira desvantagem enunciada prende-se com a ideia de que o seguro só irá cobrir a responsabilidade na extensão do envolvimento do VA, excluindo responsabilidade pessoal, bem como os valores acima da soma associada ao seguro.
- (2) Em segundo lugar, fala-se da possibilidade de existirem falsos positivos, ou falsos pedidos de ressarcimento, visto que, à partida a avaliação e verificação do cumprimento dos requisitos para reembolso será menos exigente do que quando se opta pelo regime civil de responsabilidade, que obriga a ir muito mais ao detalhe e tem um regime probatório mais forte.
- (3) O terceiro ponto vai para a falta de incentivo que este modelo poderia introduzir tanto no caso dos consumidores como dos produtores, embora não esteja cientificamente documentado e provado que,

havendo seguro, as pessoas tenham a tendência para serem mais descuidadas, ou que esse seguro tenha um impacto negativo no comportamento das pessoas.

- (4) A quarta e última desvantagem apontada prende-se com os valores do ressarcimento dos danos, que será tendencialmente inferior ao que se poderia obter no caso de um sistema baseado apenas em responsabilidade civil.

Voltaremos a esta análise na parte da discussão sobre a responsabilidade legal, onde esperamos interligar com os conceitos de Ciência Cognitiva e falar da utilidade dos regimes enunciados. Mas, para já, podemos dizer que é nosso entendimento que nenhum destes regimes, isoladamente, consegue responder às questões colocadas pela UE (nomeada e principalmente no que diz respeito ao alcance e limitação de responsabilidade, significado de defeito e ónus da prova) pelo que é de considerar um regime que aproveite as melhores características e vantagens de cada um dos sistemas já conhecidos.

3. Breves considerações Éticas

Para terminar este capítulo das considerações gerais, é essencial falarmos das questões éticas associadas aos VAs, por ser dos temas mais debatidos. Note-se que aqui não nos interessa definir que tipo de padrões éticos devem estar associados às decisões do VA, mas apenas descrever, de forma breve, este problema e os seus contornos.

Esta é uma questão que se prende maioritariamente com as escolhas levadas a cabo pelo VA, mas também pela utilização geral da IA. Algumas utilizações devem ser colocadas em causa, como é o caso da utilização militar de *Drones*, ou a simples utilização descuidada perto dos aeroportos. Numa outra perspetiva, há que considerar a utilização de *softwares* inteligentes para manipulação de opinião pública, através de notícias ou publicidades, com consequências nefastas a nível eleitoral (veja-se as últimas eleições Norte-Americanas) ou de referendos (como no caso do *Brexit*), e em última análise, com efeitos gravíssimos na democracia e liberdade.

A programação ética é uma área que está em ascensão e que pretende dotar os Agentes autónomos de capacidade para tomar decisões morais (Pereira e Saptawijaya, 2016). Relaciona-se com a análise da natureza da tecnologia, do seu impacto social e a formulação de descritivos normativos e códigos de conduta que determinam, do ponto de vista da ética, os termos do desenvolvimento da tecnologia (Moor, 1985). É uma componente importante da robótica porque os Agentes são cada vez mais sofisticados e autónomos, ao mesmo tempo que têm como objetivo final a convivência adequada com os seres humanos. Independentemente das áreas, quer seja saúde, quer seja no domínio militar ou doméstico, há valores morais que têm que ser passados para permitir uma interação normal com o homem. Existem, por isto, diversos códigos de ética formulados pelos Governos e outros tipos de entidades (como a *Association for Computer Machinery*) que espelham, ao longo dos inúmeros documentos existentes, a ideia de que o programador deve evitar o mal, no sentido de proteger a sociedade de consequências negativas, especialmente se significativas e/ou injustas. Claro está que, alguns destes conceitos, como a ideia do não praticar o mal, são tidos como algo abstratos.

É neste contexto da ética que surge a ideia do *Trolley Problem*. Creditado a Philippa Foot, filósofa que, entre outras coisas, alegou que as escolhas morais têm uma base racional, a primeira formulação deste problema (que inclui já incontáveis versões, em inúmeras áreas diferentes) assume os seguintes contornos: suponhamos que há um veículo sobre carris, cujos travões não funcionam, logo não é possível pará-lo. Este veículo dirige-se a um grupo de cinco pessoas que não podem escapar dessa colisão. Surge a possibilidade de se virar à direita, uma vez que há uma bifurcação nos carris, basta apenas que se puxe uma alavanca; mas há, nesse caminho à direita, um homem que está a trabalhar nos carris. Sem travões, somos forçados a deixar o veículo seguir o seu caminho, matando as cinco pessoas adiante, ou, puxando a alavanca, o veículo segue caminho diverso, matando uma pessoa. Que dizem as leis da moralidade sobre isto? Qual é a predefinição moralmente correta? Devemos deixar o veículo seguir o seu curso ou devemos tomar uma decisão ativa e puxar a alavanca?

O debate surge por haver, de forma global, a defesa da ideia de que os VAs serão confrontados com este tipo de escolhas na sua condução. Mas, a serem, a escolha que o VA fará pode colidir com o Código de Conduta, quando se diz que da programação não devem advir consequências negativas.

A verdade é que própria conceção do Robô, que inclui a escrita do código, pode ser usada para influenciar e, até mesmo, determinar de forma concreta, o comportamento da máquina, por forma a garantir que a mesma cumpre as normas sociais e legais do ambiente no qual se encontra (Lucivero, 2014), pelo que os programadores podem ver-se forçados a fazer esta escolha ética de antemão. De acordo com Lin (2016), os programadores terão que dar, previamente, instruções aos VAs sobre como agir em todo um leque de situações previsíveis, e conceder-lhes princípios pelos quais se deverão guiar em situações inesperadas e que não se consigam, para já, prever.

O problema apresentado não se coloca, pelo menos nos mesmos termos, no caso do ser humano, visto que, no imediato e no momento em que o indivíduo percebe que vai colidir, não existe espaço para uma decisão deliberada e ponderada, nem muito menos para reunir a informação necessária sobre o ambiente envolvente para tomar essa mesma decisão (Gogoll e Muller, 2016). Assim, no caso dos seres humanos, consentimos na escolha instintiva, qualquer que ela seja, e assumimos que não poderia ter sido tomada de outro modo.

O VA, por seu lado, não possui a capacidade de agir por instinto. Há sempre uma decisão que é tomada. Mesmo programando uma *decisão instintiva* no VA, que é possível, alguém vai fazer essa decisão de antemão. Além disso, é esperado que um computador que é capaz de se conduzir sozinho nos mais variados contextos de trânsito, nos quais processa uma enorme quantidade de informação, seja capaz de tomar, mais do que uma decisão, a decisão mais adequada, ainda que a janela temporal para escolher uma opção seja extremamente curta. E, ainda, assume-se que os VAs terão uma espécie de decisão padrão nestas situações (Gogoll e Muller, 2016). O problema aqui reside no facto de estas ações deixarem de ser acidentais porque foram determinadas previamente.

E, mais, sendo definidas previamente este tipo de escolhas, importa saber por que ética optar. Deontológica, em que regras pré-determinadas devem ser seguidas em toda e qualquer decisão? Ou Utilitarista, que visa o maior bem-estar possível, para o maior número de pessoas possível?

Uma nota para as várias teorias que têm vindo a ser enunciadas como as mais relevantes para a programação ética do VA. Tem-se falado, bastante do Kantianismo, Smithianismo, Princípios Deontológicos e Utilitarismo. As duas últimas são, em computação, as mais fortes (Goodall, 2014).

A ética Deontológica prende-se com os princípios e limites que são descritos no código da máquina como algo que não pode ser violado (é o caso das regras de Asimov). O maior problema da escolha desta visão relaciona-se com o facto de a utilização destas regras estar intimamente conectada com a nossa intuição do que é o senso comum, que correspondem, por vezes, a padrões variáveis.

O Utilitarismo, por sua vez, nas ideias inicialmente descritas por Rawls, implica a maximização do bem-estar, dadas as circunstâncias do caso concreto. É um método facilmente computável, constituindo uma vantagem nesse sentido. O problema desta programação é que, implicando uma previsão dos danos, pode produzir resultados injustos, visto que faria o sistema optar por colidir com um ciclista com capacete, em prol de um que não o use, por exemplo.

Há autores (tais como Sandberg e Bradshaw, 2013), que propõem que o proprietário do veículo escolha, ele mesmo, as predefinições éticas do seu VA.

Em resposta e refutando esta ideia, Lin (2016) lembra que tal definição pode vir a ter consequências discriminatórias (e completamente contrárias aos códigos de ética já proclamados⁵), quer em razão da etnia, quer em razão da idade, ou outra qualquer razão qualitativa ou mesmo quantitativa. Note-se que, moralmente, uma vida vale tanto quanto cinco vidas.

Adita o autor que obrigar o proprietário a escolher a ética-padrão do seu veículo pode colocar um peso demasiado grande sobre o indivíduo. Mas, assim sendo, quem decide o padrão? O autor propõe, então, que sejam os produtores dos carros ou os governos. Gogoll e Muller (2016) sugerem, neste seguimento, padrões de ética pré-definidos obrigatórios, impostos a todos os proprietários ou produtores que desenvolvem os veículos.

Outra solução apontada por Lin (2016) é a escolha arbitrária de um caminho, sem prejudicar nenhuma das pessoas envolvidas. Claro está que esta proposta apresenta inúmeros problemas, nomeadamente o facto de estarmos a decidir sobre vidas humanas de forma aleatória.

Goodall (2014) enunciou e refutou todos os argumentos a favor da ideia de que os VA não irão precisar de programação ética. Enunciamos, aqui, os mais relevantes:

⁵ Ver, a este propósito, por exemplo, *The Europeans Commission's Draft Ethics Guidelines for Trustworthy AI* ou *IEEE code of ethics*, acessível em <http://www.ieee.org/about/corporate/governance/prop7-8.html> (2014)).

- a. Os VA nunca irão colidir com nenhum obstáculo:

Segundo o autor, este argumento não faz sentido, dado que não só qualquer tecnologia pode vir a demonstrar-se imperfeita (podemos ter problemas de falhas do *hardware*, *bugs* de *software*, erros de perceção dos sensores, por exemplo), como mesmo considerando os VA como perfeitos, estão sujeitos a um ambiente com múltiplos Agentes humanos, falíveis e imprevisíveis. Finalmente, mesmo no caso em que a estrada já só tem VAs, e exista comunicação entre eles, continuam a existir animais, peões e condições atmosféricas com os quais os carros autónomos terão que lidar.

- b. Colisões que obriguem a uma decisão ética serão inexistentes:

A utilização, como exemplo, de situações extremas e hipotéticas, como é o caso do *Trolley Problem*, não invalida que decisões éticas não tenham que ser tomadas. Basta o exemplo simples de uma criança se atirar para a estrada sem que nada o fizesse prever.

- c. VAs nunca serão responsáveis por acidentes ou colisões:

Independentemente de quem é a culpa, o VA deve ter uma programação que garanta um comportamento ético e que proteja não só os seus ocupantes como as pessoas que, de algum modo, têm culpa.

- d. VAs nunca irão colidir com outros VAs:

Não podemos simplesmente assumir que os VAs serão comercializados apenas quando toda a estrada for destinada para carros autónomos. As tecnologias, principalmente as mais complexas, demoram a entrar no mercado, e estima-se que demore cerca de trinta anos até que 90% dos veículos atuais sejam substituídos por autónomos (Fraichard, 2014).

- e. Nos níveis 2 e 3 da tabela de autonomia da SAE Global, o ser humano vai estar disponível para assumir controlo, sendo ele, neste caso, o responsável pela decisão ética:

Para além do acima descrito sobre a capacidade de reação do ser humano, o autor refere que para tal seria necessário que o condutor humano, que está como que de reserva, estivesse constantemente a prestar atenção à estrada, sendo que ainda deverão ser feitos estudos no sentido de apurar o nível de atenção dos condutores nestas situações.

- f. O VA pode ser programado para obedecer à lei e, assim sendo, todos os dilemas éticos estão cobertos:

A este propósito, Lin (2016) dá o exemplo de uma árvore caída na estrada, alegando que, caso não viessem carros na direção contrária, o ser humano não teria qualquer problema em ultrapassar um traço contínuo para seguir caminho; o VA não o faria. Ficaria à espera que o tronco fosse retirado. Podem ser adicionadas leis para o VA nestes casos, mas isto implica sermos capazes de prever toda e qualquer situação que fuja do que é habitual na estrada; e, na estrada, dada a quantidade de Agentes e variáveis, poucas são as coisas que se podem considerar usuais.

- g. O VA está programado para minimizar os riscos a todo o tempo:

Este argumento pressupõe uma programação com base no Utilitarismo, que, como visto anteriormente, é apenas uma das teorias éticas das quais nos podemos servir para programar o VA.

- h. De forma global, os benefícios dos VA pesam mais do que qualquer veículo não ético:

O autor refere-se a este argumento como o mais forte, dado que qualquer atraso, dito, injustificado, em relação a uma tecnologia que, em princípio, trará tantos benefícios quanto os que se julga que os VA tragam, irá afetar pessoas de forma negativa (estamos a falar, inclusivamente, das mortes na estrada).

Embora, por enquanto, os VAs não se apresentem como menos falíveis que o ser humano, é expectável que tal venha a acontecer, ainda que não se conheçam, em concreto, os benefícios que os VAs podem vir a apresentar.

Por outro lado, esta é uma visão puramente Utilitarista, no sentido em que nos damos por satisfeitos por salvar o maior número de pessoas.

Da nossa parte, gostaríamos apenas de deixar algumas palavras de reflexão em relação a este tema. Conforme desenvolvido acima, no capítulo sobre o funcionamento dos VAs, um dos objetivos mais centrais do VA é evitar a colisão com obstáculos. Neste sentido, é sempre este objetivo final que estará presente nas decisões quando o VA é confrontado com este tipo de escolhas. O veículo tomará sempre a decisão que, na sua perspetiva, cumprirá, dentro do possível, a meta de não colidir. Significa isto que os resultados deste tipo de dilemas podem ser ainda mais imprevisíveis do que julgamos, dado que, numa situação idoso vs criança, por exemplo, dada a natureza do VA, ele não opta em função nem de uma nem de outra mas em função de um caminho que ele acha que cumpre a função de não chocar, mesmo que, no final das contas, não consiga evitar a colisão com um indivíduo, ou com o outro, ou nem mesmo com os dois.

Introdução à Agentividade

Todos nós temos uma ideia intuitiva do conceito de Agente, como algo que é capaz de agir e executar ações, por oposição a eventos que simplesmente lhe acontecem. Neste sentido básico, programas de computador podem ser considerados Agentes (Chopra e White, 2011).

De forma global, podemos dizer que um Agente é uma entidade, aqui, abstratamente considerada, que age de forma continuada e autónoma no tempo, num ambiente dinâmico, onde existem outros processos e outros Agentes (Coelho, 2008).

Nos termos anteriormente referidos, um dos pontos de partida desta dissertação seria aprofundar o conceito de Agentividade para percebermos em que medida devemos considerar, ou não, VAs como Agentes. Esta conclusão terá impactos na escolha de um regime legal, com todas as devidas consequências que daí decorrem. Se, para nós, o VA for o equivalente a um objeto inanimado, tal qual como os veículos atuais, então faz sentido utilizar os regimes atuais de responsabilidade pelo risco, porque pressupõem o controlo de quem utiliza o objeto. Se, em alternativa, olharmos para o VA como um Agente próprio, teremos a tendência para considerar o atual regime de responsabilidade dos veículos como insuficiente ou inadequado para esta tecnologia.

A ideia inicialmente pensada para esta dissertação implicaria olhar para as definições de Agentividade propostas em áreas como a Filosofia, o Direito e a Computação. Contudo, após uma investigação profunda, chegaríamos à mesma conclusão a que Shardlow, na sua tese de Mestrado de Ciência Cognitiva (intitulada *Action and Agency in Cognitive Science*, 1995), já havia chegado: dificilmente encontraremos uma definição uniforme, em áreas tão distintas. Só na filosofia, as teorias que propõem uma definição são inúmeras e correríamos o risco de nos perdemos no tema, sem conseguir optar por uma definição e, consequentemente, sem chegar a uma conclusão sobre a Agentividade no caso do VA.

Restava, assim, na melhor das hipóteses, tentar descobrir um pressuposto, que estivesse presente em todas as áreas em causa, isto é, um denominador comum, sem o qual nem sequer se discute se o ente tem ou não Agentividade.

E esse mesmo ponto em comum existe, ainda que disfarçado e talvez até mesmo com um nome ou, se quisermos, uma etiqueta diferente, em cada área. Falamos do ato voluntário. É, precisamente, este conceito que nos acompanhará a partir daqui, primeiro expondo a sua presença em cada ideia de Agentividade, na Computação, no Direito e na Filosofia. Depois, sim, discutiremos se este elemento basta e se a IA apresenta ou não voluntariedade nos seus atos.

A ideia subjacente é a seguinte: uma entidade que não seja capaz de tomar uma iniciativa, levar a cabo um ato por sua própria vontade, nunca será considerada um Agente (por exemplo, uma garrafa, sendo inanimada, não agindo senão por força de outros Agentes, nunca entrará nesta categoria). Significa isto que, agir voluntariamente, nas três áreas diferentes - Computação, Filosofia e Direito - é condição necessária, ainda que possa não ser a única, para considerarmos uma certa entidade um Agente.

1. Agentividade na Filosofia

Em filosofia este conceito demonstra-se amplamente complexo. Ainda assim, através de ideias que nasceram nas teorias da ação, conseguimos delinear um esboço do conceito de Agentividade na filosofia.

Num sentido estrito, Agentividade é vista como a execução de ações intencionais, ideia que remonta a Hume e Aristóteles. Mais recentemente, Anscombe (1957), Davidson (1963) e mesmo Lowe (2000) defenderam esta ideia, pesem as nuances entre as variadas concepções, não obstante as ideias basilares serem partilhadas. Chama-se, a esta teoria, *Standard Conception*, onde a ação intencional⁶ está intimamente ligada com a ideia de se agir por uma razão. Esta seria a razão pela qual os programas de computador, conforme sugerido por Chopra e White (2011), não são, usualmente, designados como Agentes.

Nesta teoria, não se exige causalidade entre o movimento corporal do Agente e a ação que ocorre, dado que se considera que a ação pode ser explicada pelos objetivos do Agente. Desta forma estão abrangidas ações desejadas e indesejadas, que derivam das primeiras, sem intenção, como veremos adiante no exemplo que daremos sobre Asimov.

Em conclusão, nesta vertente da teoria da ação, uma entidade é um Agente se tem a capacidade declarada de exercer intencionalidade, independentemente de esta intencionalidade dar origem a comportamentos desejados ou indesejados.

A par com esta teoria, surge também a *Standard Theory*, onde, em adição à intencionalidade nasce a noção de que é preciso, para uma ação ser intencional e racional, ser causada por estados mentais adequados e uma decorrência dos eventos na ordem correta, pelo que, por vezes, se referem a esta teoria como Teoria Causal da Agentividade. Neste seguimento, Davidson defende que ações são movimentos corporais, causados (nesta mesma ordem certa) por crenças, desejos e outros tipos de estados mentais (Shardlow, 1995).

Contudo, assumir uma decorrência dos acontecimentos na ordem certa pressupõe uma causalidade não desviante, tarefa que promete não ser fácil. Em termos abstratos, e tanto quanto conseguimos apurar, parece sempre possível que os estados mentais, que justificam uma ação, a causem por um erro, um desvio, na causalidade. A dificuldade encontra-se, precisamente, aqui, na medida em que não há como determinar a cronologia correta de eventos, em consideração a certos estados mentais, para se determinar o que é ou não desviante. Algumas soluções foram propostas (Peacocke, 1979; Bishop, 1989; Mele, 2003; Schlosser, 2007), mas não nos iremos alongar sobre este tema nesta dissertação.

Para concluir, segundo a *Standard Theory of Action* (cujos proponentes incluem Davidson, 1963, 1971; Goldman, 1970; Bratman, 1987; Dretske, 1988; Bishop, 1989; Mele, 1992, 2003; Enç, 2003), uma entidade é um Agente, na medida em que tem a organização funcional correta, isto é, a transformação de desejos, crenças e intenções em ações, segundo estes, adequadas, criando, conforme já referido, o seguimento correto de ações.

⁶ Concretamente neste contexto, intencionalidade enquanto referente a estados intencionais.

Também como indicado anteriormente, esta teoria apresenta uma lacuna séria no que diz respeito à completude da noção de sequência causal correta de acontecimentos.

Aqui é, portanto, obrigatório a determinação das razões para agir do Agente, que incluem desejos, crenças e intenções que dão origem à ação correspondente. Então, segundo esta lógica, se a IA deve ser vista como um Agente, deverá ser possível olhar para as suas ações enquanto consequências de causas iniciais, que serão as suas crenças e desejos (Chopra e White, 2011).

A diferença entre as duas teorias reside nesta ideia de ser, ou não, necessária uma cronologia dos eventos com uma ordem muito específica, que vão desde os primeiros pensamentos sobre agir, até à ação propriamente dita. Imaginemos que Asimov está em casa, à noite, com todas as luzes desligadas, a tentar adormecer. Entretanto, sente sede e decide ir buscar um copo de água. Uma vez que está escuro e precisa de luz para alcançar o copo, percebe que terá que ligar a luz para conseguir beber água. Fá-lo, com a intenção de conseguir chegar ao copo. Contudo, não se apercebeu que, do lado de fora da casa, havia um ladrão que estava de vigia, esperando o melhor momento para assaltar a casa. Ao ligar a luz, além de o fazer para alcançar o copo de água, Asimov levou a cabo um outro movimento, embora não quisesse fazê-lo: alertou o ladrão, indicando a sua presença em casa. Do ponto de vista da *Standard Conception*, o facto de o Asimov ter alertado o ladrão seria uma ação, dado que basta agir-se por uma razão. Na *Standard Theory*, mais restrita, este alerta não seria uma ação, visto que os estados mentais não foram os adequados para gerar este alerta, mas sim para beber água.

Neste domínio, embora respondendo a uma questão diversa, também se coloca a questão de se saber se Agentividade implica ou não a capacidade da entidade ser capaz de iniciar ação, ligando-se, deste modo, à ideia de causalidade. Discute-se, em filosofia, se aqui se inserem ações decorrentes de desejos e crenças (Goldman, 1970; Davidson, 1971; Dretske, 1988), embora essa ideia nem sempre seja aceite, havendo autores que defendem que não se resume apenas a tal, isto é, um Agente será também a entidade que, espontaneamente, e sem intenção prévia, desencadeia uma ação, consequentemente negando também a ideia de causa-efeito entre desejo e ação. Na prática, o que está em causa é saber se para ser um Agente a entidade tem que ser capaz de espontaneamente iniciar uma ação. Por exemplo, o livro que tomba em cima de uma chávena de café e por sua vez a derruba, por causa de um movimento humano descuidado que o atirou ao chão, não é um Agente porque não tombou sozinho, não iniciou uma ação sem mais. Ao invés, um ser humano, este, sim, um Agente, e que inicia, de facto, ações, o fez cair em primeira linha.

Embora a *Standard Theory* seja relativamente consensual na filosofia, no que diz respeito à teoria da ação, resta saber se é aplicável a entidades não humanas, visto que parece ser demasiado exigente, uma vez que explicar uma ação com base em crenças e desejos é equivalente a explicá-la com recurso a representações mentais (assume-se que cada estado intencional têm um conteúdo representacional, isto é, que são sobre algo associado a um símbolo mental, representando um objeto ou uma situação), que nem sempre é claro se entidades não humanas o fazem.

Ainda assim, e mesmo assumindo que algumas entidades não possam ser vistas como Agentes segundo a *Standard Theory*, é inegável que certas entidades, mais simples, como, por exemplo, as bactérias, aparentam ser dignas de Agentividade sem que, no entanto, possamos afirmar que o fazem através de representações mentais.

Daniel Dennet defende que se pode dizer, legitimamente, que há Agentividade quando, através de estados mentais, é possível prever, corretamente, o comportamento de uma entidade. Allen e Bekoff (1997) vieram, então, alargar esta ideia para Agentes não-humanos, dizendo que podemos aplicar esta mesma ideia a outro tipo de entidades. Por outro lado, Davidson defendeu, que só o poderemos fazer se estiver em causa uma competência linguística, como forma de transpor estados mentais.

Ainda neste âmbito, Barandiaran et al. (2009) descreveu os pressupostos para um tipo de Agentividade que se poderia considerar mínima, para estes casos em que as entidades são demasiado simples para possuir representações mentais, mas com grau de complexidade suficiente para se considerarem Agentes. Na sua opinião, os pressupostos de Agentividade, nos casos em que não há representação mental, são três:

- a. Individualidade – no sentido de haver uma distinção clara entre o Agente e o ambiente. O autor alega que os Robôs só podem ser descritos como Agentes nesta medida, visto que a sua estrutura mecânica mantém os seus componentes como um corpo unitário e porque a máquina age de acordo com alguns critérios que o seu criador entende como úteis ou coerentes.
- b. Assimetria Interacional – este pressuposto inclui o intercâmbio de energia ou matéria, dado que o sistema está ligado ao ambiente que o rodeia. Um Agente é um sistema que faz algo, por oposição a entidades naturais às quais não podemos atribuir ações. Outra perspetiva desta ideia é que o Agente tem que ser uma fonte de algum tipo de atividade, não basta ser um mero ente passivo que simplesmente sofre o efeito de forças externas. O Agente é responsável por gerir e reunir a energia necessária para agir. Barandiaran ressalva, contudo, que ser fonte de atividade não implica uma ação constante, sendo importante também a capacidade, quando pertinente, de aproveitar a influência e força da sequência dinâmica de eventos, permitindo retirar essa energia e otimizar as suas ações em termos de esforço.
- c. Normatividade – aqui, o que importa é que os Agentes tenham objetivos e normas de acordo com as quais orientam a sua ação. A normatividade tem que ser considerada um pressuposto, sob pena de se vir a incluir movimentos corporais indesejado como ações, nomeadamente, os espasmos de uma pessoa que sofre de Parkinson.

Para terminar, o autor dá uma definição de Agente: Agente é uma organização autónoma, adaptável e capaz de regular o seu emparelhamento com o ambiente e de acordo com normas estabelecidas pelas condições que permitem a sua viabilidade.

Ainda assim, é opinião do autor que esta definição não se aplica aos Robôs nem à Inteligência Artificial, muito embora isto seja, para nós, uma ideia de difícil aceitação, principalmente na medida em que o autor não oferece qualquer justificação para esta afirmação.

Na filosofia, a ideia mais presente é que, para se ser um Agente, tem que existir um sistema intencional. Também se discute as questões causais na definição de Agente, as quais defendemos, dado que, em Direito, se aplicam Teorias de Causalidade Adequada⁷ para averiguar a existência ou não de nexo de causalidade entre determinado Agente e determinado facto ilícito.

Se notarmos com atenção, classificar um Agente como aquele que age, independentemente de outras coisas que se possam adicionar a esta ideia, é uma condição bastante intuitiva e muito presente na filosofia: a ação intencional pressupõe-se como voluntária. Desde Barandiaran a Davison e Dennet, é um elemento sempre comum.

Se certa entidade não age, não porque não queira, mas porque não tenha essa capacidade, descartamo-la logo como possível Agente. Aquilo que acolhemos da filosofia é a presença de crenças e desejos que orientam a ação voluntária.

2. Agentividade na Computação

Para Minsky (1967), um Agente pode ser considerado como uma máquina de transição de estados, uma espécie de autómato de estados finito (*Finite State Machine* ou FSM), mergulhado num ambiente distribuído, caracterizado por uma distância geométrica e lógica entre Agentes (Coelho, 2008).

Uma definição um pouco menos redutora, dada por Russell and Norvig, no clássico *AI handbook* (1995, p. 33) descreve que um Agente é uma entidade que analisa o ambiente em que se encontra e age nesse mesmo ambiente através de efetores. Esta é, sem grande margem para dúvida, das definições de Agente mais utilizadas na Computação. E, como todas as definições criadas no âmbito da Computação e Engenharia são discutidas à imagem e semelhança dos computadores e IA, naturalmente (ou artificialmente) que o VA configura um Agente nos termos desta curta definição. É importante referir que o conceito de ação, sem constrangimentos, aqui, tal como na Filosofia, é uma ideia importante.

Wooldridge e Jennings (1995), por seu lado, distinguiram, no seu trabalho, *Intelligent Agents: Theory and Practice*, dois tipos de Agentividade, a saber:

- a. Fraca, que de um modo geral é utilizada para descrever um *software* que demonstra propriedades como:
 - I. Autonomia - no sentido de atuação sem intervenção direta de seres humanos, exigindo-se, aqui, um qualquer tipo de controlo sobre as suas próprias ações e estados internos (Castelfranchi, 1995);
 - II. Características Sociais - implicando que sejam capazes de interagir, incluindo através de uma linguagem comum (Genesereth and Ketchpel, 1994);

⁷ Segundo esta Teoria, consagrada no art. 10º do Código Penal Português, um evento é a causa de um resultado quando, em abstrato, se apresenta, de uma forma geral e de acordo com o seguimento normal da vida, adequado a produzi-lo (Prata, 2008).

- III. Reactividade - que denota a capacidade de o Agente perceber o ambiente no qual se insere e responder em conformidade e no momento certo às ocorrências nesse espaço;
- IV. Proatividade - que se relaciona com a tomada de iniciativa, para se atingir um objetivo proposto, não bastando a mera reação ao ecossistema.

É neste sentido que, também, Coelho (2008) descreve um Agente inteligente como um programa que está situado num determinado ambiente, é autônomo, (isto é, não é controlado externamente), reativo (responde a mudanças no ambiente), pró-ativo (persegue os seus objetivos com persistência), flexível (tem vários modos de atingir os objetivos), robusto (é capaz de recuperar após uma falha), e social (interage com outros Agentes), enquanto acrescenta noções de flexibilidade, robustez e sociabilidade. O autor desenvolve, alegando que a característica crucial de um Agente inteligente é ser pró-ativo (e, não apenas reativo, para sobreviver).

Tanto Reactividade como a ideia de Proatividade podem constituir a ideia de ato voluntário que identificamos como pressuposto de Agentividade, visto que tanto uma característica como a outra implicam que o Robô aja sem que tenha uma instrução para tal.

- b. Forte (frequentemente utilizado na Inteligência Artificial), que se trata de um sistema que soma, às faculdades acima enumeradas, outro tipo de propriedades mais complexas e que são, de uma forma geral, mais próximas da cognição humana. Neste âmbito, é legítimo a utilização de expressões como conhecimento, crença, intenção e obrigação (Shoham, 1993). Aqui, há uma defesa da ideia dos Agentes computacionais enquanto Sistemas Intencionais, segundo estes autores, na esteira de Daniel Dennet. Há, então, uma ideia de que algo importa para os Agentes, que têm crenças e desejos, numa lógica consistente e original (Taylor, 1985, 98).

Neste caso concreto, um Sistema Intencional é aquele cujo comportamento pode ser previsto (ainda que através da chamada Psicologia Popular⁸) se conhecidas as crenças, desejos por detrás desse mesmo comportamento (Dennett, 1987, p49). Aqui, Dennet distingue entre sistemas de primeira ordem e de segunda ordem. Nos primeiros, não encontramos crenças nem desejos sobre as crenças e desejos mais imediatos (ou seja, não existe a chamada Metacognição - sentimentos nem reflexões sobre desejos e crenças mais superficiais), mas nos segundos sim. Note-se que a teoria de Dennet prevê aplicação quer em sistemas artificiais como naturais (Shardlow, 1995), visto que podem ser atribuídos predicados proposicionais a entidades não-humanas (Dennett, 1987,1991).

Importa saber se é justo, legítimo ou, no mínimo, útil, atribuir esta ideia de crenças e desejos à IA.

McCarthy dá uma resposta, em 1978, mencionada no artigo que temos vindo a utilizar para discutir o tema da Agentividade na Computação, que nos apraz bastante, para além de nos ajudar a fundamentar alguns pontos sobre VAs.

⁸ Por Psicologia Popular entende-se a teoria intuitiva que o ser humano tem e através da qual explica os comportamentos dos seus semelhantes.

O autor declara que atribuir estes estados mentais é legítimo quando o que está em causa, nesses estados, é a expressão da mesma informação que seria expressada em relação a uma pessoa, mas sobre uma máquina.

Por outro lado, é importante referir que a necessidade de uma teoria sobre intencionalidade surge com o aumento da complexidade cognitiva; regra geral não usamos esse conceito para justificar o comportamento de um caranguejo (Shardlow, 1995). No caso da Inteligência Artificial a sua complexidade pode exigir a utilização desta teoria. No fundo, um Agente é um sistema sobre o qual entendemos útil atribuir crenças e desejos (Shardlow, 1995).

McCarthy afirma também que é útil na medida em que nos ajuda a compreender a estrutura da máquina em si, bem como o seu comportamento e as melhores maneiras de o corrigir ou melhorar. Para todos os efeitos, as tarefas aqui enunciadas tornam-se mais fáceis se tivermos nomes para dar às atribuições. Por outro lado, é naturalmente lógico que o façamos. Na presença de mecanismos mais abstratos e familiares para descrever o funcionamento das máquinas, torna-se mais fácil lidar com elas (Woodridge e Jennings, 1995).

Da nossa parte resta-nos saber se é justo fazê-lo, mas a essa questão só o tempo e o direito poderão respondê-lo e para isso serão precisos muitos anos.

Muito embora saibamos e conheçamos a estrutura do VA, é de argumentar que a sua decisão ocorre entre variáveis que, embora justificáveis, à luz da Teoria dos Sistemas Intencionais, desconhecemos e somos alheios, pelo que atribuir crenças e desejos, neste caso concreto, ainda que de um ponto de vista artificial, nos é útil.

Para sintetizar, um sistema será, então, considerado intencional, se puderem ter descrições como “X vai empurrar a porta se quiser sair para o exterior” ou “X executou a ação A porque acreditou que A iria resultar em maior benefício para si” em relação a esse sistema, e se essas descrições se demonstrarem explicativas, interpretativas e preditivas relativamente ao sistema e ao seu comportamento. Utilizamos este tipo de descrições, já hoje, em relação à IA, porque são úteis e não implicam um conhecimento específico dos estados internos do Agente, em relação ao qual podemos ter acesso limitado (Chopra e White, 2011). Esta visão, relativa à utilização do conceito de intencionalidade em relação à IA, não requer uma replicação perfeita do aparato biológico ou mental, pelo menos enquanto não esteja cientificamente demonstrado que intencionalidade é uma característica exclusiva do cérebro de organismos supostamente intencionais (Resnick, 1994).

Deste capítulo se conclui que a atribuição de estados mentais intencionais à IA é útil, bem como a existência de elementos, tais como a reatividade e a proatividade, que nos levam a pensar que a IA age voluntariamente.

3. Agentividade no Direito

Há quatro conceitos que serão utilizados neste capítulo, que importa, *a priori*, definir para que não se confundam: pessoa jurídica, agente, personalidade jurídica e capacidade jurídica. Dúvidas não existem quanto ao

facto de os dois primeiros serem, amiúde, utilizados como equivalentes. Ainda assim, e para que a utilização destas expressões seja bastante clara, deixam-se as seguintes definições:

- a. Pessoa jurídica - é o ente a que a lei atribui personalidade jurídica;
- b. Agente - além de significar aquele que representa uma qualquer pessoa coletiva, em direito utiliza-se este termo para designar o autor do ato ilícito culposo, gerador de danos que fica constituído na obrigação de indemnizar o lesado. No global desta dissertação, ainda assim, utilizamos a expressão Agente com um significado bem mais lato;
- c. Personalidade Jurídica - é a suscetibilidade de ser titular de direitos e obrigações, reconhecida a toda a pessoa humana (pessoa singular) e a organizações de pessoas e/ou de bens (pessoa coletiva);
- d. Capacidade Jurídica - Temos, aqui, dois possíveis significados: i) a medida de direitos e obrigações que uma pessoa pode ser titular ou ii) a possibilidade que uma pessoa tem de praticar, pessoal e livremente, atos jurídicos (exercer direitos e cumprir deveres).

A definição de Agente, na área do Direito, é bastante simples e não acarreta grandes dificuldades; é o autor do ato ilícito. Um dos requisitos para responsabilidade civil de um Agente em relação ao lesado, é, precisamente, o ato voluntário, pelo que aqui não temos qualquer dificuldade. O VA pode ser considerado um Agente, no sentido em que ele será capaz de praticar atos voluntários e na medida em que se prova o nexo de causalidade entre esse ato e os danos causados.

Contudo, este significado, à partida fácil, não é o que procuramos. Isto porque ser Agente em sede de responsabilidade civil pressupõe que o Agente é pessoa jurídica. Por outras palavras, uma definição que se afigurava de fácil aplicação ao VA, complexifica-se porque o VA para ser considerado um Agente tem que ter, primeiro, personalidade jurídica - que não tem.

Se olharmos para legislação europeia, conseguimos identificar, no mínimo, duas situações em que Sujeito, ou Agente, adaptando à discussão, e pessoa jurídica não são o mesmo: i) associações sem personalidade jurídica, ii) Nascituros⁹ e iii) animais. Nestas situações, apesar da falta de personalidade jurídica, estas entidades são Sujeitos de Direitos. Na primeira alínea a composição destas estruturas apresenta-se como algo informal, não obstante a prossecução de objetivos comuns por um conjunto de pessoas. No segundo caso, uma entidade sem personalidade jurídica usufrui de um dos maiores princípios do Estado de Direito – princípio da dignidade humana, bem como direitos que incluem o direito à vida e à integridade física, embora não seja, em boa parte, um ser humano, mas apenas potencialmente considerado como tal. Mas em nenhuma destas situações as entidades mencionadas deixam de poder ser consideradas Agentes, ainda que não tenham personalidade jurídica. Conclui-se, pois, que os conceitos personalidade jurídica e Agente em Direito, bem com a sua relação, podem, por vezes, entrelaçar-se em zonas cinzentas e de difícil definição.

⁹ Nascituros, em sentido lato, descreve aquele que ainda não nasceu, mas que já foi concebido e se encontra no útero da mãe (nascituro *strictu sensu*) bem como aquele que ainda não foi concebido, mas já foi pensado e que, embora não tendo vida nem sequer realidade biológica, tem essa potencialidade (concepturo).

Assim sendo, o que fundamenta a atribuição deste estatuto? Importa ver os fundamentos para se atribuir, ou não, personalidade jurídica. A expressão Agente, como vimos, no Direito, tem um significado diverso do que na linguagem comum. Assim, nesta área, temos que ir mais além e procurar o equivalente. A pergunta é simples, quem é que é o destinatário do Direito, quem é o interlocutor e o sujeito de Direito? Isto é, a quem é que o Direito se aplica? À pessoa jurídica.

À partida, a IA nunca poderia ser considerada como tendo personalidade jurídica, dado que esta só é atribuída, de forma natural, aos seres humanos (Solaiman, 2017) e, de forma artificial, os requisitos estão previstos apenas para empresas. Mas como veremos, pode não ser tão linear assim, porque podem existir inúmeras razões para justificar a personalidade jurídica e algumas dessas razões justificativas podem estar preenchidas pela IA.

Os juízes do Direito Anglo-saxónico preocuparam-se com o conceito de Agente em Direito, muito antes de o darmos por garantido. Em *The Nature and the Source of the Law* (1921), editado há quase cem anos atrás, o Juiz John Gray debruçou-se sobre o tema, afirmando que, em Direito, é comum usar-se o significado de pessoa como o ser humano, mas que o seu significado legal é o sujeito de direitos e deveres. Em termos similares, o Juiz Salmond, em *Jurisprudence* (1913) defendeu que ser pessoa jurídica é ter a capacidade para fazer parte de relações jurídicas. Dewey, em *The Historic Background of Corporate Legal Personality* (1926), alega que não concedemos personalidade jurídica às coisas porque o seu comportamento é exatamente igual quer lhe atribuamos direitos e deveres ou não. Nas palavras deste autor, só faz sentido dar personalidade jurídica aos entes cujo comportamento pode ser modulado pelo Direito ou aos entes através dos quais pretendemos modular o comportamento dos seres humanos. É por esta razão que outrora se concedeu personalidade jurídica aos barcos. O objetivo não era que o barco se comportasse de modo diverso, mas direcionar o ato humano através dele. Por esta lógica, é fácil perceber que, ainda que o comportamento do VA não se considerasse modelável pelo Direito, decerto que o dos seus criadores o é. Ainda assim, note-se como, nesta teoria de que só se legisla sobre entidades em relação às quais podemos modelar o seu comportamento, está bem presente a ideia de ato voluntário - se uma entidade não age voluntariamente, não vale a pena torná-lo destinatário de uma norma. Em suma, ao longo dos séculos, atribuiu-se personalidade jurídica a quem demonstrava a capacidade de agir voluntariamente ou a coisas através das quais era possível modular o comportamento humano.

Avançando no tempo e já no Direito contemporâneo, Paolo Dario e Erica Palmerini (2012) relacionam o conceito de Agentividade legal com a ideia de dever. Nesta visão, o dever, como algo exclusivo do Agente, não pode ser separado do mesmo, nem da sua ação. A noção de dever implica um poder, na medida em que se um Agente tem o dever de fazer algo, então também tem o poder de o fazer. A não existir esse poder, não poderíamos exigir do Agente aquela ação em específico, porquanto ele simplesmente não é capaz de a levar a cabo. Poder significa ter a faculdade de fazer algo. De novo, bem presente a ideia de agir voluntariamente.

Os autores até agora enunciados enquadram-se numa conceção específica de personalidade jurídica - a visão de que são pessoas jurídicas os que carregam direitos legais (Kurki, 2016). Dentro desta visão temos a distinção entre duas perspetivas: a do interesse (Mathew Kramer e mais tarde desenvolvido por HLA Hart, Nigel Simmonds e Hillel Steiner) e a da vontade.

Na primeira, defende-se a ideia de que ter um direito é equivalente a ser beneficiário de um dever. Neste sentido, A tem um direito em relação a B, se B tiver um dever em relação a A, ou seja, ter um dever em relação a alguém significa que esse dever é no interesse da entidade ou pessoa em questão. Esta é uma visão que justifica a relação parental com os menores de idade. Kramer descreve como candidatos primeiros a serem pessoas jurídicas, através desta teoria, os seres humanos. Outros autores como Joel Feinberg mencionam, também, os animais. Em todo o caso, a senciência é sempre tida em conta como uma mais valia para as entidades em relação às quais se colocam questões de personalidade jurídica, algo que, para já, a IA não parece ter, sem prejuízo de vir a apresentar esta característica, um dia.

A segunda teoria, a da vontade, por sua vez, prende-se com a ideia de controlo dos direitos e deveres do próprio.

A visão de que são pessoas jurídicas e, portanto, Agentes, os que carregam direitos legais é a parede mestra de um processo que decorreu em Dezembro de 2014, no Supremo Tribunal de Nova Iorque, onde a representação de um chimpanzé chamado Tommy (*Nonhuman Rights Project*) pediu o alargamento do âmbito de aplicação da definição de pessoa, por forma a permitir a atribuição de direitos a um animal, neste caso, em concreto, um *habeas corpus*. Citou-se em defesa deste pedido, definições do *Black's Law Dictionary*, que descreve que pessoa jurídica é qualquer ente que a lei entende como capaz de ter direitos e deveres (Kurki, 2016).

Ora, defendendo o *Nonhuman Rights Project* que os animais conseguem e têm capacidade para carregar, pelo menos, um direito, o Projeto pediu ao tribunal que declarasse Tommy como uma pessoa jurídica.

Pietrzykowski (2017), descreve situação similar num tribunal na Argentina, com a juíza Elena Amanda Liberatori ditando que o animal, neste caso um Orangotango no Zoo de Buenos Aires, deveria ser olhado como uma pessoa jurídica não humana, com os seus próprios direitos que devem ser respeitados pelas pessoas em geral, ressalvando, naturalmente, que esta personalidade jurídica deve ser adequada às capacidades do animal sob pena de lhe conferirmos direitos e deveres inapropriados.

Muita da jurisprudência que surge é necessária pela instrumentalização dos animais na prossecução de finalidades como a proteção de interesses e direitos humanos e, logo, reconhecidos como sujeitos que devem ser protegidos pelo ordenamento jurídico. Adicionalmente, e numa outra vertente, é transparente a ideia de que os seres humanos tendem a desenvolver estados emocionais relativamente aos animais, aumentando o número de razões pelas quais tendemos a proteger estas entidades, muito embora elas não tenham a menor noção do que significa ser sujeito de Direito ou sequer o que é o Direito.

Numa outra perspetiva, Morse (2000) vê na racionalidade o critério primário para a existência de um Agente legalmente responsável, logo implicando, esta afirmação, que a sua ausência trará motivos para acreditarmos que não há Agentividade, ainda que a definição de racionalidade possa não ser o mesmo num e noutro momento ou num e noutro lugar. Morse, ainda que aceite a racionalidade como critério para uma Agentividade legal, não se compromete com nenhum tipo de definição específica de racionalidade.

Calverley (2008) vai noutro sentido, alegando que para existir personalidade jurídica é necessário existir autonomia e intencionalidade, aqui intencionalidade no sentido anteriormente descrito por Brentano e Searle (1999), de existirem pensamentos que são sobre e que representam objetos e estados do mundo¹⁰. É esta também a opinião de Chopra e White (2011), quando afirmam que a IA é um Agente intencional porque o seu comportamento pode ser submetido a uma série de generalizações empíricas, ligadas com ações intencionais. No entender destes autores, a IA pode, de forma coerente, ser considerada um Agente para efeitos legais, já que as suas ações, podem ser a causa de eventos relevantes no âmbito do Direito.

Contudo, é fácil ver que o problema não está em saber se há ou não personalidade jurídica e/ou capacidade jurídica, porque isso só depende do legislador. Sem prejuízo, claro, da permeabilidade do Direito às marcas culturais (Ferreira e Dias Pereira, 2017), que é, também, indiscutível, o legislador decide se o cão tem ou não personalidade, se a criança tem ou não personalidade e a partir de quando. O legislador conferiu personalidade e capacidade jurídica até às pessoas coletivas. Como vimos, até barcos chegaram a ter personalidade jurídica. Por isso nos importa investigar mais, entender o que leva o legislador a tomar esta decisão. O legislador concede personalidade jurídica tendo em conta condições que julga serem pertinentes para o caso¹¹. No fundo, decidir quem tem a capacidade para ter direitos e deveres é uma escolha que se baseia em considerações morais – no caso dos seres humanos - ou pragmáticas - como é o caso das empresas (Pietrzykowski, 2017). Será que razões pragmáticas podem justificar vermos o VA como um ente legal?

A razão pela qual se frisa esta ideia tem o seguinte seguimento: independentemente daquilo que possamos acreditar sobre a cognição da IA e do Robô, aquilo que a lei vier a determinar tornar-se-á crença geral, mais cedo ou mais tarde. Se a lei vier a decidir que o Robô deve ter personalidade jurídica, independentemente das razões nas quais se fundamente para o determinar, o Robô será um Agente e, mais que isso, sujeito jurídico.

Por fim, e mais importante, é notar que as várias teorias, sobre o que significa ser um Agente em Direito, na medida em que se possa atribuir personalidade jurídica, quer recorram à intencionalidade, à racionalidade ou ideia de poder/dever, pressupõem a capacidade, acima de tudo, de agir voluntariamente.

4. Conclusões

A identificação de uma ação pressupõe a identificação do seu Agente como causa de uma ação (Chopra e White, 2011).

Ao enumerarmos várias teorias que participam neste debate sobre o que significa ser um Agente, mostrámos que todas elas pressupõem a ação voluntária, preocupando-se, assumindo esta condição mínima, em justificá-la através da racionalidade, intencionalidade ou outros meios.

¹⁰ Assim, dizer que “A Casa Branca é em Washington” é um pensamento e uma afirmação sobre a Casa Branca e sobre um traço característico de Washington (Calverley, 2008).

¹¹ Esta é, certamente, uma visão discutível. Contudo, não nos cabe aqui discutir nem confrontar questões de Positivismo Jurídico com as de Direito Natural.

Assim, o que está em causa é saber se há ou não um ato voluntário, porque isso é requisito mínimo para qualquer uma das definições de Agente que enunciámos, bem como, e com especial atenção, para qualquer coisa que se faça no âmbito do Direito - para haver responsabilidade civil, para haver responsabilidade penal, para se considerar válido um negócio jurídico. É por isso que importam, e é por isso que olhámos, as várias noções de Agente, aquele que age voluntariamente, não coagido, autonomamente, com base no seu próprio racional.

O Ato Voluntário como elemento comum

A vantagem da utilização do conceito ato voluntário é que nos permite evitar debates sobre questões de autonomia. Deste modo, conseguimos pressupor que o Agente é autônomo, sem termos necessariamente que nos alongar nas discussões sobre a autonomia.

A pergunta seguinte é: será que os Robôs agem voluntariamente? Ainda que haja uma definição prévia (no código ou *software* do Robô) daquilo que o Robô acredita, a verdade é que o seu comportamento encontra justificação nessas crenças e desejos. A IA, ainda que numa dimensão fraca de Agentividade, é capaz, como vimos, de reagir ao ambiente que a rodeia, depois de o analisar e tomar decisões voluntárias, conforme o objetivo que lhe interessa. É, ainda, capaz de iniciar uma ação se assim entender como conveniente, inclusive ao ponto de surpreender os seus próprios criadores. Deste modo, reforça-se a ideia de que o VA, bem como outros tipos de Robôs, são Agentes, ainda que admitamos diferentes graus de Agentividade e ainda que pensemos no VA como um Agente mais simples.

Não será, decerto, difícil compreender que comum às ideias de Agentividade, independentemente das áreas enunciadas, está subjacente a ideia que o Agente é aquele que autonomamente, sem coação, determina para si um qualquer ato que, será, então, voluntário.

Claro que se argumenta que o racional da máquina lhe foi inculcado, mas esta é uma discussão que, assumidamente, não podemos ter aqui; se o fizéssemos, teríamos, também, que debater sobre as origens do racional humano e argumentar que, ao contrário da IA, o nosso não foi pré-determinado, discussão que se mantém há anos e, provavelmente, perdurará, ainda, por muito tempo.

É interessante perceber-se como a noção de ato voluntário não é totalmente desligada daquilo que a nossa intuição nos diz. Já em 1917, Walter Wheeler Crook, que escreveu, precisamente, sobre o ato voluntário, um artigo intitulado *Act, Intention, and motive in the Criminal Law* descreveu que, nos tempos que corriam, era habitual falar-se do crime como composto por dois elementos: (1) um ato e (2) uma intenção ou estado mental em nome do qual o ato seria feito.

Voltando ainda mais atrás no tempo, William Markby, quando escreveu *Elements of Law - Principles of Jurisprudence*, descrevia o ato voluntário como um movimento corporal que se segue após uma volição que o orienta. Isto em 1896, note-se bem, 84 anos antes da definição que acima demos de Davidson, que descreveu do mesmo exato modo o princípio basilar da Agentividade.

Mais tarde, Hart (*The Morality of the Criminal Law*, 1965) afirmou que num sistema legal cívico, apenas os que poderiam ter cumprido a lei devem ser punidos. Quer isto dizer que não deve ser punido alguém que não poderia ter evitado fazer o que fez (portanto executou um ato não voluntário) ou que não possua a capacidade de raciocinar e desenvolver estados mentais que orientem o seu comportamento no sentido de cumprir a lei. Já aqui se apresenta uma noção de voluntariedade, que hoje, no Direito moderno, é absolutamente determinante para questões de responsabilidade quer do ponto de vista do Direito Criminal como da filosofia da obrigação moral (Silber, 1967). Mais do que uma expressão no Direito Positivo, conforme nos descreve Beleza

dos Santos (1949, p.89), também a nossa formação moral nos indicia que culpar alguém por algo que não fez é injusto, demonstrando que, mais do que uma lei, é uma intuição do ser humano. O nosso ordenamento jurídico preconiza, inclusivamente, esta ideia no art. 13º do Código Penal.

No Direito, sem margem para dúvida, é expressamente proibido punir alguém por algo que não fez (com as devidas adaptações, claro está, nos casos da omissão de ação que devia ter sido praticada e não foi). Para algum tipo de castigo, tem, obrigatoriamente, que existir um ato que lhe antecede e o justifica. Punir alguém que não cometeu um ato criminoso é, por si só, absurdo, e implicaria contrariar todos os princípios basilares do Direito, desde o Princípio do Direito do Estado Democrático ao Princípio *nulla poena sine culpa* (não há pena sem culpa).

Conforme nos descreve Gideon Yaffe, em *The Voluntary Act Requirement*, para efeitos jurídicos, o ato voluntário é um movimento corporal querido, no sentido de desejado. O autor continua dizendo que se um estado mental que guia o movimento corporal não é consciente, então não podemos dizer que é um ato voluntário porquanto não existe o elemento cognitivo volitivo, isto é, a vontade. A este propósito, vale a pena recordar o exemplo sobre Asimov dado no capítulo da Filosofia, que quando liga a luz por forma a conseguir pegar no copo de água, mas acaba por alertar o ladrão à espreita. Embora os dois atos pareçam voluntários, apenas o primeiro é feito com intenção (com vontade) e em posse de todas as informações, livre de qualquer tipo de ignorância. Temos assim, dois critérios para definir ato voluntário em Direito: um movimento corporal e uma vontade que lhe dá origem. Contudo, existem mais requisitos para o ato ser voluntário.

Para Serena Olsaretti, autora na área de Filosofia, uma ação é voluntária se não é involuntária e será involuntária se não existem alternativas aceitáveis (em conformidade com um padrão objetivo, e não subjetivo nem aleatório, que pode ser, por exemplo, o bem-estar, conforme defende também Colburn, em 2008). A questão interessante é que a autora mencionada sugere que a atribuição de responsabilidade está ligada com questões de voluntariedade, algo que já temos vindo a defender anteriormente. Por outro lado, a questão da voluntariedade abarca o conceito de autonomia, também um tema muito discutido relativamente à IA, e autonomia no sentido de determinação e governação do próprio. Isto porque só uma entidade autónoma pode ter atos voluntários.

Importa saber o que são ou não alternativas aceitáveis. A autora defende que uma opção inaceitável é aquela que acarreta um dano específico para o Agente ou, em alternativa, o caso em que uma norma moral é de tal modo imperativa que as restantes escolhas são inaceitáveis.

A autora indica, também, que um ato ser voluntário ou não, depende das motivações do Agente e, consequentemente, das crenças do Agente sobre as opções à sua disposição; caso as crenças estejam, de algum modo, minadas pelo erro, o Agente pode ter várias opções à sua frente, mas julgar que não as tem. Desde modo, o ato pode ser involuntário por falta de informação. Assim sendo, para esta autora, Asimov alertar o ladrão é também um ato involuntário, pelo facto do nosso interveniente desconhecer a existência do ladrão.

Aristóteles alegou o mesmo, anos e anos antes, em *Ética a Nicómaco*, dizendo que as únicas razões para um ato não voluntário são a força (externa) e a ignorância. Ora, esta é a razão pela qual, em Direito, não se pune

alguém que prove estar coagido aquando da sua ação, nem se considera crime a morte de alguém com uma doença cardíaca, quando assustado por outro indivíduo, se este indivíduo não tivesse como saber da existência da doença do primeiro.

Mas fora estes casos, todo o restante ato será, então, voluntário. Por outras palavras, o ato é voluntário quando há um movimento corporal, uma vontade que lhe dá origem, não estando na ignorância sobre as opções nem sobre coação.

Posto isto, analisaremos um silogismo sugerido por Aristóteles, na parte III do livro acima descrito:

1. Somos moralmente responsáveis por todas as ações voluntárias;
2. As crianças e os animais não-humanos agem voluntariamente;
3. As crianças e os animais são moralmente responsáveis

Vamos retirar a palavra moralmente, visto que o que nos interessa é analisar a responsabilidade de um ponto de vista mais global, mas manter o restante do silogismo:

1. Somos responsáveis por todas as ações voluntárias;
2. As crianças e os animais não-humanos agem voluntariamente;
3. As crianças e os animais são responsáveis;

Agora, iremos simplificar a linguagem por forma a podermos garantir que o silogismo continua válido:

1. Todo o ente que age voluntariamente é responsável - Premissa Maior
2. As crianças e os animais não-humanos agem voluntariamente - Premissa Menor
3. As crianças e os animais são responsáveis - Conclusão

Neste seguimento, podemos facilmente verificar que o silogismo é formalmente válido (através das regras da lógica aplicáveis ao caso):

- a. Temos três termos: o maior (é responsável), o médio (age voluntariamente) e o menor (as crianças e os animais);
- b. O termo médio (age voluntariamente) não entra na conclusão;
- c. O termo médio (age voluntariamente) é utilizado em toda a sua extensão universal (em “todo o ente que”);
- d. Não há duas premissas negativas, logo há uma conclusão a tirar;
- e. Não há duas premissas particulares, mas sim uma geral e uma particular, logo podemos retirar uma conclusão;

Ora, a questão que se coloca é a seguinte: se o silogismo é válido, porque é que não se punem as crianças e os animais? Porque agir voluntariamente não é condição suficiente para se considerar um Agente como entidade responsável, como veremos adiante. A Inteligência Artificial, tal como as crianças e os animais, tem decisões, e consequentemente atos voluntários. Significa isso que devem ser punidos? Ou o facto de agirem voluntariamente apenas nos diz que são Agentes, mas não chega para afirmar que são Agentes legalmente responsáveis?

Com efeito, as crianças e os animais não são punidos. A razão para tal acontecer é porque a primeira premissa deste silogismo é falsa ou, pelo menos, incompleta. Não somos necessariamente responsáveis apenas porque os nossos atos são voluntários. É preciso mais. Procuraremos justificar esta ideia daqui para a frente.

Para isto, importa introduzir o conceito de inimputabilidade, aplicável a seres humanos que têm, de algum modo, a sua cognição comprometida. É usual indicar-se como inimputáveis as pessoas que sofrem de anomalia psíquica e os menores de sete anos (488º/2, Código Civil). Conforme descrito de forma simplista por Carlota Pizarro de Almeida, em *Modelos de Inimputabilidade - da teoria à prática* (pág. 21), falar de imputabilidade significa que, em circunstâncias bem determinadas pela lei, o facto, na sua correspondência jurídica, não pode ser atribuído ao autor - não há responsabilização do autor que praticou o ato, visto que não tem capacidade de culpa.

Ter capacidade de culpa implica que o Agente, no momento da prática do facto, seja capaz de atuar de forma responsável, compreendendo que o facto é ilícito e determinando-se com essa compreensão, decidindo-se à sua não realização. O Agente tem que ser capaz de refletir sobre o seu comportamento e valorá-lo positiva ou negativamente. Só se pode culpar o sujeito Agente do ilícito penal assumindo que o mesmo é detentor das faculdades psíquicas e físicas mínimas requeridas para poder motivar-se pelos mandados normativos. Ao conjunto de todas essas faculdades mínimas requeridas para considerar um sujeito como culpável por ter cometido um facto ilícito, dá-se o nome de imputabilidade (Muñoz Conde, 1996).

A existência deste instituto justifica-se com os fins das penas e com a culpa, não só num plano causal (precisamente, na sua ligação com o ato voluntário) mas ainda no sentido de reprovação da sociedade perante aquele ato (Pizarro de Almeida, pág. 22). Se por um lado, o princípio da culpa obriga a que a existência deste ato voluntário culposos seja pressuposto da pena, bem como o seu limite máximo (art. 40º/2 Código Penal), por outro, a reprovação da sociedade só é passível de surtir efeito se o seu destinatário se encontra capaz de entender essa valoração negativa que lhe é dirigida. Não desenvolvendo temas sobre os fins de prevenção geral das penas, prevenção penal ao nível macro da sociedade¹², no que diz respeito a fins especiais, ao nível micro do indivíduo, àquele que não compreende a aplicação de uma pena que lhe é imposta, estes fins ficam inutilizados, porquanto ele não poderá ser influenciado por ela (Pizarro de Almeida, pág. 28).

Quer isto dizer que, mesmo os inimputáveis, são suscetíveis de castigos; desde que sejam adaptados à sua compreensão (Pizarro de Almeida, pág. 29).

¹² Estas teorias impendem sobre a ideia de demover potenciais criminosos de cometer o mesmo crime, na medida em que vêm as suas consequências diretamente, bem como a estabilização e tutela da confiança nas normas do Direito.

Não são, no entanto, suscetíveis de penas; se não há pena sem culpa (40º/2, CP) então os inimputáveis não podem sofrer uma pena (Pizarro de Almeida, pág. 34), sem prejuízo de outras medidas e segurança.

Conforme já referimos, a inimputabilidade implica uma incapacidade global para se entender o juízo negativo que é feito pela sociedade à sua ação. Em alternativa, e nos termos do que dita o Código Penal, no art. 20º/1, será a impossibilidade de avaliar a ilicitude de um ato e de orientar a sua conduta tendo em conta essa avaliação. Isto porque a cognição da pessoa está de tal modo comprometida que, embora consiga ter crenças e desejos imediatos que o levam a agir, não consegue fazê-lo a um nível secundário, refletindo sobre essas mesmas ações e estados mentais que levam a ela.

Em Filosofia e Psicologia Cognitiva, a estes estados mentais sobre outros estados mentais, onde se inclui a capacidade de nos autocensurarmos por pensamentos que podem dar origem a comportamentos ilícitos, dá-se o nome de Metacognição.

Indiscutivelmente, podemos usar teorias filosóficas para justificar muita coisa no Direito. Já vimos que há definições que se assemelham nessas duas áreas tão distintas e muitas das bases do Direito Moderno vieram de homens como Kelsen, Hart e Austin, todos eles também filósofos, de algum modo. Mas o Direito serve-se de muitas outras áreas e acolhe conceitos da vida mundana através delas. Não seria possível legislar sobre Medicina se o Direito não se servisse dos conceitos comuns desse domínio. Morse (2003) argumentou que a lei usa modelos de ação que derivam da Psicologia Popular. Por entendermos que esta afirmação, bem como a anterior, são verdadeiras, então é legítimo usar ideias de Psicologia e Filosofia, neste caso sobre Metacognição.

Em conclusão, a ideia a reter é que é Agente aquele que age voluntariamente, mas só é Agente responsável aquele que possui Metacognição.

A Metacognição

1. A Metacognição enquanto pressuposto mínimo da imputação

Além da defesa da tese de que o ato voluntário é o elemento principal da Agentividade, tem-se vindo a defender ao longo destas páginas, que quando afirmamos que determinado Agente é responsável por uma ação, estamos, implicitamente, a dizer que o Agente deve ser capaz de oferecer uma justificação para essa ação, através das suas crenças ou desejos. Em princípio, podemos afirmar que o Agente é sujeito de todo um leque de atitudes, incluindo reações básicas como ressentimento, gratidão, censurabilidade, entre outros (Matthias, 2004). Para que um Agente possa ser considerado responsável, ele deve ter controlo sobre o seu comportamento e noção das suas consequências, sempre no devido enquadramento da situação concreta (Fischer e Ravizza 1998: 13). Isto implica, não só, não estar sobre coação, como ter a liberdade para escolher e para optar pela alternativa mais viável tendo em conta estes factos.

Frankfurt (1971) defendeu, em filosofia, que a diferença entre o ser humano e outros tipos de Agentes se encontra na estrutura da sua vontade, no sentido em que apenas o ser humano reflete sobre as suas motivações e, assim, forma estados mentais de segunda-ordem (estados cujos objetos são outros estados, estes últimos diretamente causadores de ações e objetivos). Por exemplo, se A tem que aprender uma matéria lecionada nas aulas, pode refletir sobre os vários métodos de aprendizagem que conhece e escolher um de acordo com o que julga ser mais eficaz. Temos aqui pensamentos sobre a metodologia, que por sua vez irá influenciar a aprendizagem. Aprender é um processo cognitivo. Ao refletir sobre este processo cognitivo (escolhendo uma metodologia de estudo), A está a fazer uso de estados de segunda ordem. Esta ideia de atitudes de ordem superior, que se enquadra, inevitavelmente, na Metacognição, será, segundo o autor, o que distingue a Agentividade de um ser humano da que é perpetuada por um Agente não humano. Daqui decorre uma distinção: dentro do conceito de Agente (que o é porque age voluntariamente), encontramos dois graus de Agentividade, um grau mais complexo que o outro, que envolve estados cognitivos mais elaborados, atribuído, de um modo geral, aos seres humanos, e outro grau menos complexo, atribuído a outras entidades não humanas, na medida em que lhes falta esta capacidade para desenvolver atitudes de segunda-ordem.

Verdade seja dita, o Direito acaba por punir apenas na circunstância de existir, não só uma Agentividade, mas mais, uma Agentividade complexa. Inclusivamente, para Cox (2005), qualquer Agente inteligente confrontado com a escolha de o que fazer (incluindo no caso de estar em causa a prática de um ilícito), deve decidir três coisas: (1) qual das ações é a mais adequada na situação presente, (2) se pensou o suficiente sobre essa escolha e se essa escolha é suficientemente informada ou se é preciso refletir mais sobre as escolhas (estando em causa pensamentos sobre os próprios pensamentos) e, finalmente, (3) perceber se há algo que correu mal e porque é que tal aconteceu. Estas três decisões estão diretamente ligadas com o que se chama de Metacognição, já que isto implica um pensamento crítico auto-reflexivo, ou seja, a avaliação das ideias em termos de qualidade e fazendo juízos sobre se estas fazem ou não sentido (Martinez, 2006).

A Metacognição, também descrita como a cognição sobre a cognição (Fleming, Dolan e Frith, 2012), entende-se como tendo propósitos de monitorização do comportamento e processos mentais (Nelson e Naren, 1990).

A capacidade de analisar a situação concreta, contemplar as várias opções que temos, mensurar, em cada uma, o lícito e o ilícito e fazer juízos normativos e morais, é uma forma de auto-controlo do comportamento e processos mentais. A Metacognição relaciona-se, no fundo, com a consciência e métodos de gestão dos conteúdos do próprio pensamento (Kuhn e Dean, 2004).

A literatura descreve dois componentes essenciais da Metacognição: o conhecimento sobre a cognição e a monitorização da cognição (Lai, 2011). O primeiro componente, que é o que é relevante para esta tese, relaciona-se com a noção das nossas próprias capacidades e limitações, incluindo fatores externos e internos que podem afetar a nossa cognição (Flavell, 1979). Este componente é extremamente importante no que diz respeito à definição de estratégias, porque é este o raciocínio que nos permite escolher uma estratégia em vez de outra, bem como justificarmos esta escolha, perante uma tarefa final. Voltando ao caso acima descrito da metodologia de estudo, A tem que estar ciente sobre as suas limitações e capacidades para escolher corretamente o método de estudo que mais se adequa a si e que lhe permitirá ter melhores resultados.

O primeiro ponto relevante para defender que a Metacognição é essencial no Direito reside no seguinte: qualquer pessoa que pretenda cometer um ilícito, tem que ter esta capacidade de análise estratégica. Um inimputável pode cometer um ilícito, sem dúvida. Contudo, a sua intenção foi de praticar o ato X e não de praticar um ilícito. O indivíduo que planeia um ato ilegal, por sua vez, pensa sobre o objetivo final, reflete sobre limitações internas e externas, e define uma estratégia, conhecendo a ilicitude do comportamento e motivado pelo receio de ser apanhado. Em todo este processo descrito há recurso à Metacognição. Há uma diferença entre querer praticar o ato X e X ser ilícito, como no caso do inimputável e pratica o mesmo ato X, sabendo que esse ato é ilícito. O primeiro é punido; o segundo não, precisamente porque lhe falta a capacidade para refletir sobre estratégias que lhe permitam contornar a lei e porque não existe a motivação do medo de se ser apanhado.

Para fundamentar esta ideia servir-nos-emos da chamada Teoria do Comportamento Planeado, atribuída a Ajzen (1991). Segundo este autor, e conforme melhor explicado na figura abaixo, a intenção do Agente é modulada por: (1) atitudes individuais face ao comportamento que está em causa, (2) pressão individual relativamente à conduta em específico e (3) controlo comportamental.

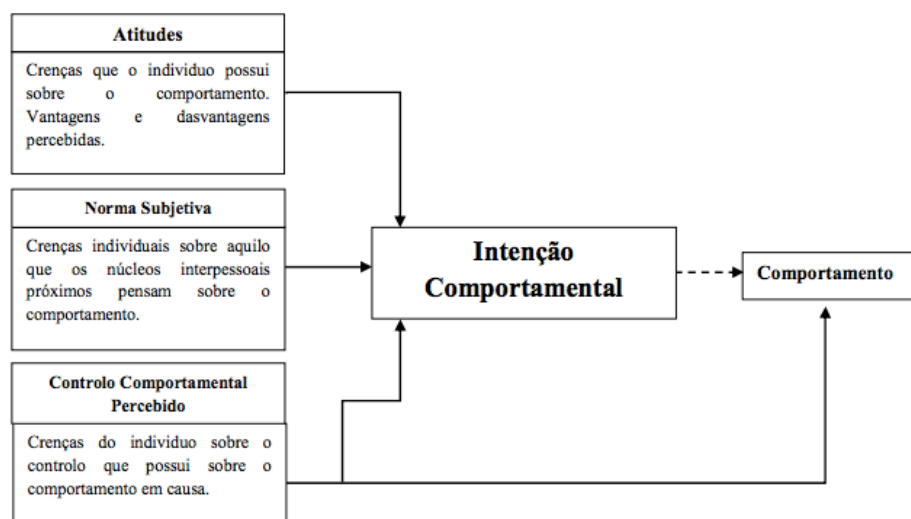


Fig. 14 - Representação esquemática da Teoria do Comportamento Planeado (Luís, 2014).

O comportamento, em si mesmo, é um evento que implica um nível de cognição superficial e direto, no sentido em que há uma relação simples de pensar-fazer. A intenção que lhe dá origem corresponde a um segundo nível de cognição, e estes três fatores que modulam, indiretamente, o comportamento, representam o terceiro e mais profundo nível de cognição.

Esta descrição sobre níveis de cognição pode parecer *ad hoc*, mas não é. Na verdade, o outro ponto relevante para a utilização da Metacognição no Direito relaciona-se com o facto de a Metacognição envolver três níveis de consciência sobre os factos. A história ou a situação atual com que o Agente se depara é o primeiro nível, seguindo-se um segundo sobre os raciocínios que o Agente faz sobre a própria história, e o terceiro nível de reflexão que é feito sobre este segundo (Fig. 15). As crianças conseguem realizar um primeiro e segundo nível, mas um terceiro já será mais complicado. No caso dos animais, se a situação com a qual se deparam for simples, conseguem entendê-la como uma história no primeiro nível, e, também aprender através do segundo. Mesmo a crença de receber um biscoito se derem a pata implica um segundo nível. O evento, a história contada, no primeiro nível, é que o homem lhe deu um biscoito, no segundo está a compreensão de que isso ocorreu porque ele deu a pata.

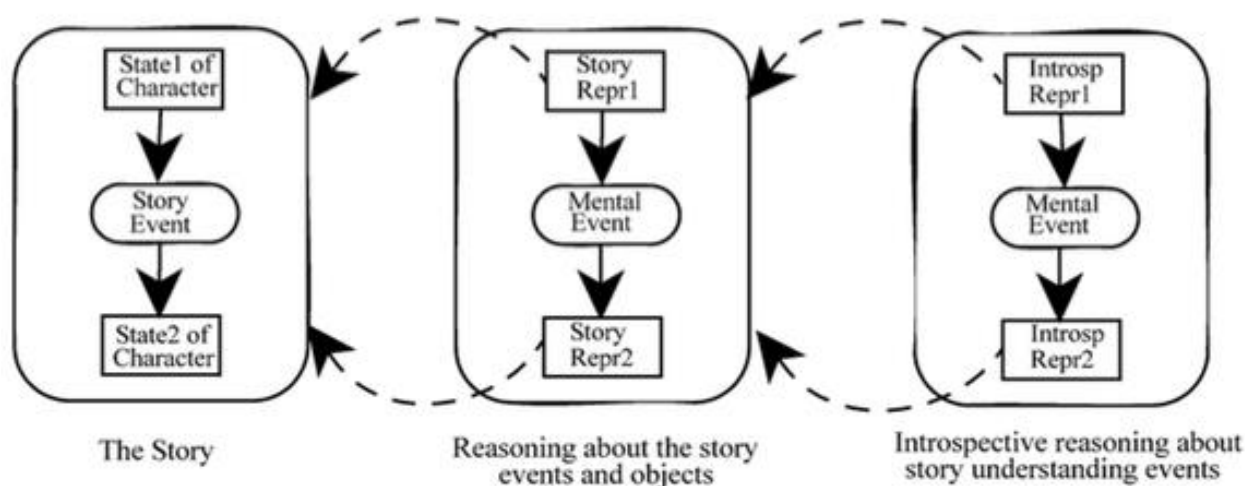


Fig. 15 - Níveis de Metacognição, de acordo com Cox (2005).

Assim, e a título de exemplo, imaginemos que HAL está numa loja e vê alguém a roubar. Cria-se todo um aparato, a polícia é chamada e o indivíduo é detido. HAL observa tudo atentamente. Esta é a situação de nível um com a qual é confrontado. HAL continua a sua vida e reflete sobre o acontecimento, bem como a ilicitude, e a punição que presenciou sobre o indivíduo que assaltou a loja; entramos no nível dois. Finalmente, HAL é também capaz de fazer um raciocínio introspetivo em relação à primeira reflexão. Por exemplo, pode ter entendido, inicialmente, ter sido injusta a punição do indivíduo, mas mais tarde sentir-se envergonhado por ter pensado nisso. Ou aperceber-se que ele não via naquele roubo um ilícito e ficar com medo sobre a sua possibilidade de agir em termos semelhantes. No fundo, estamos a falar de juízos sobre os juízos que estão na base do comportamento.

Este tipo de Metacognição que nos referimos quando falamos de incapacidade para fazer juízos sobre a ilicitude de um ato e de orientar a conduta nesses termos é autonóética (Metcalf e Son, 2012), na medida em que implica considerar o autor destes estados mentais como referente, ou sujeito, destas reflexões. É, segundo

os autores mencionados, o tipo de Metacognição que implica um desenvolvimento maior da consciência. A pessoa que possui consciência auto-noética é capaz de estar ciente do seu próprio futuro projetado, bem como do seu passado (Metcalfe e Son, 2012) e, logo, sendo capaz de se colocar numa situação eventual sobre a sua punição, quando praticados determinados atos ilícitos.

Aqui não argumentamos que as crianças não possam levar a cabo operações que impliquem Metacognição, até porque é óbvio que conseguem. A aprendizagem requer Metacognição e as crianças aprendem constantemente. Há inúmeros estudos, são descritos por Petros Georgiades (em *From the general to the situated: three decades of Metacognition*), que comprovam que as crianças são capazes de algum tipo de Metacognição. Contudo, o tipo de Metacognição que falamos é aquele que é mais exigente em termos de consciência, pelo que será de admitir que as crianças não a tenham ou, pelo menos, não suficientemente desenvolvida.

Este tipo de Metacognição que temos vindo a falar, também descrita como estratégica (Saxe e Offen, 2010), é, de facto, a que parece faltar nos casos de doença psíquica, ainda que possa variar de doença para doença (David, Bedford, Wiffen e Gilleen, 2012). O caso das crianças, ainda que diferente, é tradicionalmente tratado do mesmo modo; a reflexão sobre o próprio comportamento implica um pensamento operacional formal (Piaget, 1976) o que raramente é atribuído às crianças (Brown e Deloache, 1978). Com efeito, estudos demonstram que a Metacognição estratégica começa a desenvolver-se por volta dos 14 anos, ainda que seja incerta a idade em que está totalmente desenvolvida (Schraw e Moshman, 1995).

Pelo ora exposto, se conclui ser defensável que a Metacognição estratégica, auto-noética, enquanto pressuposto da imputabilidade é também, em si, um pressuposto mínimo para a responsabilização por um comportamento ilícito e, logo, para a atribuição de responsabilidade.

2. A Metacognição na Inteligência Artificial

De uma forma global, a Metacognição não é um tema que tenha vindo a ser muito debatido na IA. É, contudo, uma ideia presente, embora não nos termos que temos vindo a discutir no caso do ser humano. Aqui, a ideia fundamental é de conhecimento da cognição e refere-se ao que um sistema sabe sobre a sua própria cognição ou mesmo sobre a cognição em geral. Por outras palavras, conforme descrito por McCarthy, a Metacognição, na Computação, estará ligada a uma introspeção, no sentido de a máquina ser capaz de ter uma crença sobre os seus próprios estados mentais, mais do que apenas fazer considerações sobre o meio envolvente.

Outros autores desenvolvem a mesma ideia, utilizando apenas palavras diferentes, sendo a Metacognição a capacidade de o sistema de pensar sobre como e o que ele próprio pensa (Crowder e Friess, 2011).

Alegadamente, em Computação, a Metacognição é utilizada para a resolução de problemas como a seleção de algoritmos, em que o sistema deve escolher entre o algoritmo que resolve o problema apresentado de forma mais eficiente. Adicionalmente, conceitos como controlo executivo e processos de monitorização, os quais descrevemos no capítulo do funcionamento dos VA, são importantes para gerir a complexidade dos problemas e avaliar o progresso em relação aos objetivos finais (Cox, 2005).

Crowder e Friess descrevem, neste âmbito, três tipos de Metacognição:

- a. Conhecimento Metacognitivo - é o que o sistema sabe sobre si mesmo enquanto processador cognitivo (Kosko, 1986);
- b. Regulação Metacognitiva - trata-se da regulação da cognição e da aprendizagem, que pode, inclusive, ser baseado naquilo que o sistema tem noção que não sabe (laBar e Cabeza, 2006);
- c. Experiência Metacognitiva - são as experiências passadas que têm algo que ver com a missão atual do sistema (Crowder e Friess, 2011). Aqui, o próprio sistema cria expectativas do que pode acontecer dadas essas experiências passadas.

Daqui se deduz que a IA tem Metacognição, na medida em que é inserido no seu desenvolvimento. A questão aqui é a de saber se possuem o tipo de Metacognição que lhes permite analisar e valorar a sua conduta, e conformar o seu comportamento com essas ideias.

E, pelo menos aos dias de hoje, parece bastante óbvio que tal não acontece. Ainda que possamos admitir processos metacognitivos de reduzida complexidade, especialmente se estivermos a falar do significado de Metacognição no domínio da Computação, seria difícil defender que, confrontado com a situação acima descrita do HAL, um Robô fosse capaz de refletir com complexidade igual ou similar. Não sendo capaz de exercer processos cognitivos nos termos de reflexão sobre ilicitude, somos obrigados a deixar de fora este tipo de responsabilidade para os Robôs e, logo, para os VA.

Para terminar e em jeito de conclusão, vale a pena recordar o quadro da Fig. 1, novamente, do qual nos servimos para mostrar que um dos primeiros passos para determinar um regime legal era perceber que tipo de Agente temos diante nós. As três hipóteses apresentadas por Ugo Pagallo eram, então, considerar o VA uma pessoa jurídica, um Agente próprio, com um regime específico, ou uma mera fonte de dano, equiparável a um objeto.

Responsible robot	<i>Immunity</i>	<i>Strict liability</i>	<i>Unjust damages</i>
As legal person	I-1	SL-1	UD-1
As proper agent	I-2	SL-2	UD-2
As source of damage	I-3	SL-3	UD-3

Fig. 1 - Os nove tipos ideais de responsabilidade, de acordo com Ugo Pagallo.

Na medida em que o VA é capaz de agir voluntariamente, que vimos ser requisito mínimo para Agentividade, somos obrigados a excluir, assim, a ideia de que ele seria, apenas, uma fonte de dano. Iremos, deste modo, excluir, regimes que o equiparem a este estatuto, nomeadamente a responsabilidade objetiva pura e dura, sem prejuízo de utilização deste regime com as devidas adaptações.

Por outro lado, também concluímos que esta entidade, embora um Agente, não possui pelo menos um dos elementos essenciais que permita a Lei atribuir algum tipo de responsabilidade - a Metacognição. Ora, não nos sobra, então, espaço, para o considerarmos uma pessoa jurídica.

Por exclusão de partes, o estatuto do VA é o de Agente próprio, numa zona cinzenta entre o objeto e a personalidade jurídica. Deve, por esta razão, aproveitar um regime próprio, novo, criado à imagem das suas complexidades e especificidades. É isto que iremos propor no capítulo seguinte.

Discussão sobre a Responsabilidade Legal

Nos termos descritos por Figueiredo Dias, a responsabilidade é, num contexto filosófico geral e numa conceitualização jurídica, o termo mediador entre o comportamento do Agente e a consequência ou efeito que aquele acarreta para este. Este termo mediador deve ser adaptado em função do Agente que temos diante nós. Agente implica sempre ser capaz de atuar de forma voluntária. Agente responsável implica ser capaz de atuar de forma responsável e de demonstrar a utilização da Metacognição. Por conseguinte, o nosso VA pode ser um Agente, mas há que ter cautela na forma como se responsabiliza esta entidade, tendo em conta o pouco grau de desenvolvimento da sua Metacognição.

Como vimos, no capítulo anterior, bem como na análise do relatório da UE, a hipótese mais viável seria criar um instrumento totalmente novo de regulação, e não apenas copiar um regime similar ao que já existe (por exemplo, utilizando as normas legais aplicáveis às crianças ou aos animais), para abranger os novos riscos, incluindo o seguro de responsabilidade objetiva. Visto que o único regime efetivo de responsabilidade civil a ser mencionado nos documentos analisados foi a responsabilidade pelo risco (ponto b1 do 2.1., do respetivo capítulo), é expectável que a UE mantenha estes dois regimes em paralelo.

Há cinco pontos que de certo modo resumem os objetivos de um regime legal do VA, que os descreveremos de imediato. Aquilo que é preciso ter em conta na análise deste estatuto é que se apresenta como uma solução instrumental cujo objetivo é simbolizar tangivelmente o VA, por forma a colmatar o fosso de responsabilidade. Neste seguimento, tal como no caso das pessoas coletivas, o estatuto jurídico da IA seria formalizado na miscelânea de responsabilidades e relações obrigacionais que são inerentes à constituição, funcionamento e cessação de atividade da máquina. Em conformidade, por forma a obter personalidade jurídica, o Robô teria que estar registado, num registo a desenvolver nos mesmos termos que o registo comercial, só se tornando sujeito ativo e passivo de obrigações no momento do registo. Esta é uma ideia sugerida por Susanne Beck (2014).

Os cinco pontos fulcrais deste estatuto serão, então:

- a. **A existência de um Estatuto Próprio, visto que o atual não se demonstra adequado à tecnologia em questão, conforme as conclusões do relatório da UE sobre o tema; este novo estatuto deve responder às questões: (1) *Software* pode ser considerado produto? (2) Sendo, quando é que este produto se pode considerar defeituoso? (3) Quais os limites da responsabilidade do fabricante? (4) Até que ponto os novos riscos tecnológicos podem ser atribuídos à indústria (produtores não são responsáveis por riscos que não se poderiam saber, objetivamente, ou que eram de algum modo indetectáveis – *bugs*, riscos de design - nem por danos subsequentes)?**

Apesar de alguns autores, como Garza (2011), defenderem que a legislação atual consegue responder de forma assertiva aos desafios da tecnologia, a UE defende, e nós concordamos, que a não existir uma alteração na legislação atual, não se resolvem os problemas enunciados no ponto 3.2. desta dissertação. No mínimo, a legislação atual precisa de ser completada. Como vimos, o VA apresenta um estatuto para o qual a lei ainda

não está preparada, entre objeto e pessoa jurídica e também isto fundamenta a ideia de que a legislação atual precisa de ser alterada.

É necessário conhecer o que é que se entende por produto. É preciso criar critérios objetivos para determinar o que se insere ou não neste estatuto, esclarecendo se os VA podem ou não pertencer a esta categoria e, em caso negativo, levar a cabo os esforços necessários para ampliar o conceito jurídico de produto. O VA, não obstante, ser um Agente autónomo, continua a ser um produto na medida em que é fabricado por seres humanos e comercializado com fins lucrativos, integrado num mercado. Por outras palavras, o facto de ser um Agente próprio não invalida que seja um produto.

É também necessário definir, neste estatuto, o que é um defeito. Se um VA causa um dano a um terceiro por causa de uma reação sua, dificilmente podemos afirmar, sem mais, que se trata de um defeito, porque, provavelmente, a única coisa que o VA fez foi reagir a novos *inputs* e adaptar o seu comportamento em conformidade (Beck, 2015). Ainda assim, as maiores preocupações são ao nível da segurança. Aqui defende-se que o padrão de segurança deve ser, no mínimo, o do condutor humano. Se assim não for, estaremos a trazer para as estradas piores condutores do que o ser humano, quando o maior argumento para a utilização desta tecnologia é, precisamente, a segurança. E, claro está, as consequências podem ser nefastas. Em março de 2018 assistimos à primeira morte por um VA, com o modelo da Tesla a não ter detetado um transeunte. Se não exigirmos o mais alto padrão de segurança, estaremos a compactuar com este tipo de acidentes.

O fabricante deve ser responsável pelos riscos que conhece ou devia ter conhecido, quando coloca a tecnologia em circulação. Até aqui nada de novo, visto que é o que acontece com qualquer produto atualmente. A responsabilidade pelo produto, tão antiga quanto o Direito Romano, pretende compensar os que de algum modo sofreram um dano pela negligência do fabricante operando *ex post* (Colonna, 2012), e aqui não nos parece que devam existir alterações.

Deve, contudo, haver um limite a esta responsabilidade. Neste momento, os meios tradicionais de atribuição de responsabilidade estão comprometidos, no caso em concreto, já que vão contra os nossos sentidos intuitivos de justiça e de moralidade, na medida em que ninguém tem controlo suficiente sobre a máquina para que nos pareça minimamente justo atribuir responsabilidade a essa pessoa (Mattias, 2004).

Há autores que defendem que quanto maior a capacidade de aprendizagem do Robô, menor será a responsabilidade do fabricante (Beck, 2015). Esta ideia pode fornecer pistas para o limite da responsabilidade do produtor que pode, talvez, ser avaliado de acordo com critérios que se relacionam com a prova do nexo de causalidade entre o acidente e as escolhas dolosas ou negligentes do fabricante. Quando a hegemonia dos criadores dos VA se perde, e os meios utilizados pelo VA já só são definidos pelo ser humano de forma indireta, a melhor maneira de entender as decisões do sistema é fazendo correr o seu próprio código (Grimmelmann, 2005), justificando as suas próprias escolhas (Chopra e White, 2011).

Caso não se consiga provar esse nexo, entramos na responsabilidade do proprietário, pelo mero facto de ser dono do carro.

Os custos técnicos desta prova, que inclui a utilização de peritos, ainda devem ser determinados. Mas existe menor esforço probatório caso seja o fabricante com o ónus da prova do que o proprietário do VA, visto que estão em posse de muito mais informação.

b. A obrigatoriedade de um seguro de responsabilidade objetiva (independente de culpa), nos termos mencionados pela EU.

Este ponto é exigível para facilitar uma harmonização a nível europeu, visto que alterar a legislação teria um grau de complexidade acrescido, bem como para garantir uma salvaguarda mínima e um patamar mínimo de ressarcimento dos danos, seja em que situação seja. Nenhum VA poderá operar na ausência de um seguro. Esta medida também facilita e torna mais céleres, em larga medida, todos os processos de compensação de danos por acidente.

c. A manutenção paralela de um regime de responsabilidade civil que, pelo estudo cognitivo feito em relação ao VA, bem como a assunção que a tecnologia será cada vez mais desenvolvida, não deverá constituir um regime de responsabilidade pelo risco, conforme sugerido pela UE, mas sim algo mais semelhante à responsabilidade dos pais em relação às crianças ou dos donos em relação aos animais; aqui podemos acolher a visão de Ugo Pagallo sobre a utilização do regime romano dos escravos e a reanimação da figura jurídica do *Peculium*.

O maior argumento para se manter os dois regimes em paralelo é a incerteza da tecnologia e do seu impacto e funcionamento real.

O regime mais adequado para competir com o seguro obrigatório (quer porque se entendeu simplesmente disponibilizar estes dois meios de ressarcimento de danos, quer porque o lesado pretende obter uma indemnização maior do que a permitida pelo seu seguro) será uma responsabilidade equivalente ao dos pais com as crianças, os donos com os animais¹³, os representantes dos maiores acompanhados.

As crianças e os maiores acompanhados são, muitas vezes, descritas como um ente que possui uma *quasi* personalidade jurídica e é frequente fazer-se esta comparação com os Robôs (veja-se, Asaro, 2007 e Beck, 2015).

As máquinas atuais já estão aptas para decidir sem qualquer intervenção humana. As regras segundo as quais agem não são, necessariamente, fixadas *ab initio*, aquando do seu desenvolvimento, podendo vir a ser alteradas a cada momento de funcionamento do sistema, pelo próprio (Mattias, 2004).

Isto porque, tal como no caso das crianças ou dos animais, o VA é capaz de ter atos voluntários, como vimos ao longo da dissertação, mas tal como enunciado, também, falham no que significa ter Metacognição. E, como

¹³ Duffy e Hopkins, 2014; Cerka, Grigiene e Sirbikyte, 2015; Solaiman, 2017; Pietrzykowski, 2017; Adams, 2010 para mencionar alguns autores.

vimos, em Direito este acaba por ser um pressuposto mínimo de imputação. Em resultado, eles serão responsáveis pelos próprios atos, mas não legalmente responsáveis e, muito menos, puníveis por eles.

É importante fazer-se uma ressalva neste domínio. Alguns autores, tal como Chopra e White (2011) defendem que deve ser aplicado o regime da representação legal com a IA e, a fazer sentido esta aplicação, poderíamos sugerir também a utilização da ideia da relação de Comitente-Comissário. Até porque os VA irão, autonomamente, agir no interesse do seu proprietário. Não parece, contudo, ser uma solução aceitável, uma vez que, tanto no regime da representação legal como na relação Comitente-Comissário (art. 500º C.C.), as orientações dadas a quem age por conta de outrem, relativas aos objetivos da missão atribuída, bem como métodos e outra informação relevante, são fornecidas pela pessoa em nome de quem se age. Toda as regras a seguir são dadas por este. No caso dos VAs, a única coisa que o proprietário decide, quando o VA conduz em seu nome, é o destino. Tudo o resto não é definido por ele. Ora, no regime atual da Representação e Comissariado existem escapatórias que permitem proteger o sujeito representado, quando as ações dos que agem em seu nome ultrapassam as orientações que ele deu. No caso dos VAs, seria extremamente difícil encontrar uma maneira de proteger o proprietário, que não indicou mais do que o destino a chegar.

Compará-los com elevadores (King, 2017; Zohn, 2013) ou veículos normais (Zohn, 2013), é demasiado redutor para a tecnologia em questão, e não permite acompanhar a complexidade cada vez maior da mesma. Por outro lado, seria inútil atribuir a responsabilidade diretamente ao VA, visto que não teríamos como o punir de forma direta. Ele não tem património sequer nem reage cognitivamente à aplicação de uma multa. A única maneira de o fazermos seria através do proprietário.

Não havendo falha mecânica nem computacional, entramos na responsabilidade do VA propriamente dito, através do respetivo proprietário, que poderá ser responsável por negligência, caso tivesse a obrigação de tomar certos cuidados e não o fez, por exemplo, manutenção (Kalra, Anderson e Wachs, 2009). Em alternativa, tendo feito tudo ao seu alcance para evitar o acidente ou provando que mesmo que tivesse feito tudo ao seu alcance, o acidente ocorreria na mesma, podemos pensar em recorrer à figura do *peculium*¹⁴ para ressarcir estes danos¹⁵, conforme se utilizava com os escravos no Direito Romano (Pagallo, 2013). Aqui deverá haver uma compensação máxima associada e o lesado fica mais desprotegido neste caso específico, mas outro modo não seria justo na medida em que o VA toma decisões sem qualquer mão do proprietário, portanto há um limite de responsabilidade neste caso.

Contudo, dadas as dificuldades que pode acarretar para o fabricante a verificação do código cada vez que há um acidente, ele poderá gozar de uma espécie de proteção, no sentido de só se poder ir procurar a responsabilidade no Código que a empresa escreveu quando em nenhum outro lado a encontrámos. Isto evita que o

¹⁴ Em termos gerais, o *peculium* era um património que o chefe da família deixava ao seu escravo para que ele pudesse administrá-lo, sob as ordens deste mesmo chefe de família, sem que, ainda assim, este dinheiro ou estes bens passassem para a esfera patrimonial deste escravo.

¹⁵ Inclusivamente para eventualmente resolver o problema das falhas indetetáveis, que podem surgir pela complexidade da tecnologia (Vladeck, 2014)

fabricante tenha constantemente e a cada acidente, o ónus de fazer um *trace* das decisões do código para descobrir se tiveram ou não impacto no acidente.

d. A distribuição das responsabilidades por entidades externas sempre que possível (entre o peão que é descuidado, o fabricante que foi negligente ou doloso no desenvolvimento do VA e o proprietário que não leva o VA à revisão anualmente obrigatória, por exemplo).

Este é um ponto importante do estatuto (e já havia sido sugerido por Beck, em 2015) apenas porque permite evitar, ao máximo, o vazio na responsabilidade ou o chamado *Responsability Gap* (Matthias, 2004 e Asaro, 2007) quando se culpa o VA na sua condução. Em última análise, a responsabilidade deve ser atribuída a algo ou alguém identificável (Bivins, 2016). Trata-se de atribuir a culpa à mão humana na medida do possível, visto que, para já, o maior problema continua a ser como punir a IA (Wilks, 1985, 279). Desta forma o sistema legal só terá mais complicações na medida em que nenhum erro humano explique o acidente.

Importa referir que esta ideia resolve tanto a responsabilidade no caso de ainda haver colaboração entre ser humano-VA na condução, como a situação em que apenas é o VA a conduzir, sozinho. E, aceitando isto, podemos servir-nos em quase tudo de conceitos legais já existentes.

Por outras palavras, na existência de um fosso de responsabilidade, o mais interessante, ainda que complexo, seria distribuir as responsabilidades o mais possível. Sempre que possível atribuir a outra causa que não o VA, a saber: o fabricante, o programador ou o peão descuidado que se atira na passadeira. Caso contrário, atribuída uma responsabilidade pelo risco, o proprietário será sempre culpado por toda e qualquer coisa que suceda, seja ou não culpa dele. É que antes, o peão atravessa-se na estrada, sem olhar, mas é o condutor que tem o volante nas mãos, para todos os efeitos, ainda que não tivesse os reflexos suficientes para parar, assume-se que existe algum tipo de poder nas suas mãos, além da vantagem que colhe com a utilização do carro. Com um VA, o proprietário nem sequer tem o volante nas mãos; não haverá rigorosamente nada que ele possa fazer.

e. A possibilidade de conjugar responsabilidades.

Importa que este estatuto preveja a responsabilidade solidária de vários Agentes. Um mau algoritmo de reconhecimento de objetos pode estar na causa do acidente, mas é possível que haja outras contribuições como, por exemplo, uma câmara que não funciona em condições (Asaro, 2007). Poderá ser aplicável, com contornos especiais, no caso de vários módulos defeituosos que compõem o veículo (Vladeck, 2014).

Em termos gerais, o Direito vive na ideia base de que quem colhe uma vantagem, como seja usufruir de uma casa que lhe pertence, é também responsável pelas desvantagens que, de algum modo, lhe sejam apresentadas. Assim sendo, ao nosso hipotético proprietário da casa mencionada, caberia todo o tipo de obrigações inerentes ao direito de propriedade; seria o exemplo da reparação de uma torneira ou de uma tomada elétrica.

Não seria justo, contudo, que na eventualidade de um dano causado na esfera do proprietário, por outra pessoa, ao invés de ter sido causado pelo mesmo ou por mero acaso fortuito, não se garantissem as proteções

adequadas ao ressarcimento deste dano, uma vez que o dano ocorreu por mão de outrem, prejudicando o direito do indivíduo a quem pertence a casa. Porque o Direito existe, na dimensão e em nome da Justiça, estas proteções existem também, consagrando a reparação da desvantagem criada, sempre na linha e medida daquilo que é possível remendar.

A responsabilidade pelos danos provocados na esfera jurídica de outrem alicerça-se em determinados princípios e objetivos. Sempre com o objetivo final de reparar um mal feito a outra pessoa, essa reparação não poderá, em caso algum, ultrapassar o limite da ofensa, diga-se de passagem, a ofensa efetivamente e não apenas virtualmente, causada.

Por outro lado, cumpre referir que o instituto da Responsabilidade Civil se baseia num ilícito. Para as situações que se nos afiguram relevantes, o ilícito a que nos referimos trata-se daquele que é resultado de uma atuação contra uma situação jurídica alheia.

Há um caso em que esta figura não se baseia num ilícito, caso esse que também nos importa olhar. Falamos da Responsabilidade Civil pelo risco. Não está, aqui, em causa, qualquer tipo de comportamento censurável, mas apenas a atuação que colhe uma vantagem de uma determinada atividade, que por sua vez comporta um risco elevado de causar um dano. Esta atividade não é ilegal, mas ainda assim conhece limitações à sua utilização, onde se insere esta ideia da Responsabilidade Civil pelo risco que a atuação, de forma natural, emana.

Conclusão

O conceito de personalidade jurídica anda de mãos dadas com a ideia de Agentividade (Asaro, 2007). Ser um Agente, como vimos, implica diferentes condições, conforme a teoria e o domínio de debate.

Ainda assim, o ato voluntário é o elemento presente em praticamente todas as definições de Agentividade, nas três áreas de estudo desta dissertação. Deste modo, mesmo que queiramos deduzir que o VA é um Agente, notamos que o mais importante seria perceber se os seus atos são, ou não, voluntários; a definição de Agente é maleável, portanto chegamos a um beco sem saída; a de ato voluntário, por outro lado, não é, e é requisito mínimo para a existência de Agentividade. O VA é, assim, um Agente, porquanto age voluntariamente.

Posto isto, vimos que a ideia de ato voluntário está intimamente ligada com a atribuição de responsabilidade. Contudo esta teoria não justifica porque não atribuímos responsabilidade às crianças, aos animais, nem às pessoas com doenças psíquicas, na devida medida em que agem voluntariamente. A voluntariedade, só por si, não justifica a atribuição de responsabilidade a um Agente.

Concluimos, analisando a realidade jurídica, que não punimos crianças nem animais porque pressupomos que a Metacognição está de algum modo comprometida ou não totalmente desenvolvida, isto é, este tipo de entes não consegue ter reflexões sobre os seus estados mentais primários.

Vimos também que a Metacognição estará, pelo menos aos dias de hoje, ausente nos VA, muito embora, como referido, o mesmo não se possa dizer dos atos voluntários.

Parece, então, haver, no âmbito do Direito, mais do que um Agente, na vida real, que age voluntariamente, mas não tem capacidade de ter cognição sobre a sua própria cognição. Mas não deixam de ser Agentes por isso, porque continuam a ser capazes de agir voluntariamente. Os VA são Robôs que sabemos que irão interagir com o ambiente e com os seres humanos e que tomam decisões voluntárias, sem constrangimento humano, que têm relevância legal. Pode não ser suficiente para lhe atribuímos outros tipos de Agentividade, nem personalidade jurídica como a conhecemos, mas definitivamente lhes deve ser concedido um estatuto de Agente próprio, do ponto de vista legal (Floridi e Sanders, 2004).

Neste sentido, entende-se que os VA se enquadram na categoria de Agente próprio, conforme descrito por Ugo Pagallo. Os VA são aqui vistos como detentores de uma *quasi*-personalidade (Asaro, 2007). Determinar que esta é a categoria do VA, implica dar-lhe um regime próprio, adequado às necessidades da comunidade que dele se vai servir e tendo em conta as suas limitações a todos os níveis.

Isto implica, sem dúvida e pela atividade desempenhada pelos VA:

- a. A existência de um Estatuto Próprio, visto que o atual regime não se demonstra adequado à tecnologia em questão, e está bastante disperso conforme as conclusões do relatório da UE sobre o tema;

- b. A obrigatoriedade de um seguro de responsabilidade objetiva (independente de culpa), nos termos mencionados pela UE;
- c. A manutenção paralela de um regime de responsabilidade civil que, pelo estudo cognitivo feito em relação ao VA, bem como a assunção que a tecnologia será cada vez mais desenvolvida, não deverá constituir um regime de responsabilidade pelo risco, conforme sugerido pela UE, mas sim algo mais semelhante à responsabilidade dos pais em relação às crianças ou dos donos em relação aos animais (pelas razões enunciadas sobre o ato voluntário e a metacognição); aqui podemos acolher a visão de Ugo Pagallo sobre a utilização do regime romano dos escravos e a reanimação da figura jurídica do *Peculium*;
- d. A distribuição das responsabilidades por entidades externas sempre que possível (entre o peão que é descuidado, o fabricante que foi negligente ou doloso no desenvolvimento do VA e o proprietário que não leva o VA à revisão anualmente obrigatória, por exemplo);
- e. A possibilidade de conjugar responsabilidades.

Em todo o caso, a conclusão final desta dissertação é que o VA deve ser considerado um Agente, mas não um Agente punível, por lhe faltar o requisito mínimo de responsabilidade. Deste modo, é desejável um estatuto jurídico próprio e, tanto quanto possível, flexível, que seja capaz de acompanhar a evolução tecnológica e no qual se possa abarcar o maior número de situações possível, dada a incerteza dessa evolução.

Referências

- ❖ Adams, Wendy. (2010), Human Subjects and Animal Objects: Animals as 'Other' in Law, Journal of Animal Law and Ethics, Vol. 3, obtido através de <http://ssrn.com/abstract=1632624>.
- ❖ Ajzen, I. (1991), *The theory of planned behavior*, Organizational Behavior and Human Decision Processes, Volume 50, Issue 2, Pages 179-211, DOI: [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T).
- ❖ Allen, C. and M. Bekoff. (1997), *Species of mind: The philosophy and biology of cognitive ethology*, MIT Press, Cambridge, Massachusetts, ISBN: 0262511088 9780262511087.
- ❖ Anderson, M. & Anderson, S. (2007), Machine Ethics: Creating an Ethical Intelligent Agent, Ai Magazine, 28, 15-26, DOI: <https://doi.org/10.1609/aimag.v28i4.2065>.
- ❖ Andrews, Kristin (2014), *The Animal Mind: An Introduction to the Philosophy of Animal Cognition*, Routledge, obtido através de https://www.researchgate.net/publication/263066889_The_Animal_Mind.
- ❖ Anscombe, G. E. M. (1957), *Intention*, Harvard University Press, ISBN 0674003993 9780674003996
- ❖ Asaro, P.M. (2007), Robots and responsibility from a legal perspective, Proceedings of the IEEE, 20-24, 2007, obtido através de <http://peterasaro.org/writing/ASARO%20Legal%20Perspective.pdf>.
- ❖ Asvadi, A., Premebida, C., Peixoto, P. e Nunes, U.J. (2016), 3D Lidar-based Static and Moving Obstacle Detection in Driving Environments: an Approach Based on Voxels and Multi-region Ground Planes, Robotics and Autonomous Systems, ELSEVIER, 2016, DOI: 10.1016/j.robot.2016.06.007.
- ❖ Aristóteles, *Ética a Nicómaco*, tradução de António Castro Caeiro (2004), 5ª Edição, Quetzal, Lisboa.
- ❖ Barandiaran, X., Di Paolo, E. & Rohde, M. (2009), Defining Agency. individuality, normativity, asymmetry and spatio-temporality in action (v. 1.0.) Journal of Adaptive Behavior, in press. (Rohde, M. & Ikegami, T. (Eds) Special Issue on Agency.), obtido através de http://barandiaran.net/textos/defining_agency.
- ❖ Beck, S. (2016), The problem of ascribing legal responsibility in the case of robots, AI & Soc, DOI 10.1007/s00146-015-0624-5.
- ❖ Beleza dos Santos, J. (1949), *Lições de Direito Criminal*, Coimbra, Almedina.
- ❖ P. Birks, Definition and Division: A Mediation on Institutes, 3.13, in THE CLASSIFICATION OF OBLIGATIONS 1 (1997), ISBN-10 0198265980.
- ❖ Bishop, John (1990), *Natural Agency: An Essay on the Causal Theory of Action*, Cambridge University Press, ISBN 9780521063975 9780521374309.
- ❖ Bivins, T.H., (2006), Responsibility and accountability, In *Ethics in public relations: responsible advocacy* (eds K Fitzpatrick, C Bronstein), pp. 19–38. London, UK: Sage Publishing, DOI: <http://dx.doi.org/10.4135/9781452204208.n2>.
- ❖ Boulangé, A., Jaggie, C. (2014), Ethique, responsabilité et statut juridique du robot compagnon : revue et perspectives, Cognition, Affects et Interaction, obtido através de https://www.researchgate.net/publication/278625871_Cognition_Affects_et_Interaction.
- ❖ E. Bratman, Michael, (1987), *Intention, Plans and Practical Reason*, Bibliovault OAI Repository, the University of Chicago Press, DOI: 10.2307/2185304.
- ❖ Brown, A. L., & DeLoache, J. S. (1978), Skills, plans, and self-regulation, In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 3-35). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc, obtido através de https://www.researchgate.net/publication/49176582_Skills_plans_and_self-regulation.
- ❖ Calverley, D. J. (2008), Imagining a non-biological machine as a legal person, AI & Soc, 22:523–537 DOI: 10.1007/s00146-007-0092-7co.
- ❖ Castelfranchi C. (1995) Guarantees for autonomy in cognitive agent architecture. In: Wooldridge M.J., Jennings N.R. (eds) *Intelligent Agents. ATAL 1994. Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence)*, vol 890. Springer, Berlin, Heidelberg, DOI: https://doi.org/10.1007/3-540-58855-8_3.
- ❖ Churchland, Pat & Pievani, Telmo & B.M. de Waal, Frans & Parmigiani, Stefano. (2014). Evolved morality: The biology and philosophy of human conscience. *Behaviour*. 151. 137-141. 10.1163/1568539X-00003158.
- ❖ Cohen, J. (2018, 22 de Maio), Sensor Fusion, obtido através de <https://medium.com/@jeremyscohen/sensor-fusion-90135614fde6>, acedido a 10 de Outubro de 2018.
- ❖ Colburn, B. (2008), Debate: The Concept of Voluntariness, Journal of Political Philosophy, 16(1), pp. 101-111, DOI: <https://doi.org/10.1111/j.1467-9760.2007.00272.x>.
- ❖ Cox, M. T. (2015), Metacognition in computation: A selected research review, Elsevier, Volume 169, Issue 2, December 2005, Pages 104-141, DOI: <https://doi.org/10.1016/j.artint.2005.10.009>.
- ❖ Crowder, J. A., "Machine Learning: Intuition (Concept) Learning in Hybrid Genetic/Fuzzy/Neural Systems." NSA Technical Paper CON_0013_2003_009 (2003).
- ❖ Crowder, James & Friess, Shelli & Ncc, Ma. (2011), Metacognition and Metamemory Concepts for AI Systems, obtido através de <https://pdfs.semanticscholar.org/40a6/f4e061af3d9f7ee3480c6be7a21379207e7f.pdf>.
- ❖ Cook, W.W., Act, Intention, and Motive in the Criminal Law, Source: The Yale Law Journal, Vol. 26, No. 8 (Jun, 1917), pp. 645-663, Published by: The Yale Law Journal Company, Inc., Stable URL: <https://www.jstor.org/stable/786267>, Accessed: 19-09-2018 06:48 UTC.

- ❖ Čerka, P., Grigienė, J., Sirbikytė, G. (2015), Liability for damages caused by artificial intelligence, *Computer Law & Security Review*, Volume 31, Issue 3, Pages 376-389, DOI: <https://doi.org/10.1016/j.clsr.2015.03.008>.
- ❖ Chopra, S., & White, L. (2011), *A Legal Theory for Autonomous Artificial Agents*, Ann Arbor: University of Michigan Press. doi:10.3998/mpub.356801.
- ❖ Coelho, H., *Teoria da Agência: Arquiteturas e Cenografia*, Edição do Autor, 2008.
- ❖ Colonna, Kyle (Fall 2013), Autonomous Cars and Tort Liability, *Case Western Reserve Journal of Law, Technology & the Internet*, Vol. 4, No. 4, 2012. Available at SSRN: <https://ssrn.com/abstract=2325879> or <http://dx.doi.org/10.2139/ssrn.2325879>.
- ❖ Czubenko, M., Kowalczyk, Z. & Ordys (2015), Autonomous Driver Based on an Intelligent System of Decision-Making A, *Cogn Comput*, 7: 569, <https://doi.org/10.1007/s12559-015-9320-5>.
- ❖ David, A. S., Bedford, N., Wiffen, B., Gilleen, J. (2012), Failures of metacognition and lack of insight in neuropsychiatric disorders, *Philos Trans R Soc Lond B Biol Sci*, 367(1594): 1379–1390, DOI: 10.1098/rstb.2012.0002.
- ❖ Davidson, D. (1963), Actions, reasons, and causes, *Journal of Philosophy*, 60 (23):685-700, DOI: 10.2307/2023177.
- ❖ Dennett, D.C., 1987, *The Intentional Stance*, Cambridge, MA: MIT Press.
- ❖ Deng L., Yu, D., 2014, Deep Learning: Methods and Applications, *Foundations and Trends in Signal Processing*, Volume 7 Issue 3–4, June 2014, Pages 197-387, DOI: 10.1561/20000000039.
- ❖ Dewey, J. (1926), The Historic Background of Corporate Legal Personality, *Yale Law Journal*, Vol. XXXV, No 6, obtido através de http://rci.rutgers.edu/~tripmcc/phil/dewey-historic_background_of_corporate_legal_personality.pdf.
- ❖ Dretske, F., (1988), *Explaining Behavior: Reasons in a World of Causes*, Cambridge, MA: MIT Press, DOI: 10.2307/2185177.
- ❖ Duffy, S., Hopkins, J. P. (2013), Sit, Stay, Drive: The Future of Autonomous Car Liability, 16 *SMU Sci. & Tech. Law Rev.* 101, disponível em http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2379697.
- ❖ Dwivedi, P. (2017, 30 de Abril), Tracking a self-driving car with high precision, obtido através de <https://towardsdatascience.com/helping-a-self-driving-car-localize-itself-88705f419e4a>, acessado a 10 de Outubro de 2018.
- ❖ Enç, B., (2003) *How We Act: Causes, Reasons, and Intentions*, Oxford: Oxford University Press, DOI:10.1093/0199256020.001.0001.
- ❖ Fagnant, D. J. & Kockelman, K. M. (2013), Preparing a Nation for Autonomous Vehicles: Opportunities, Barriers and Policy Recommendations for Capitalizing on Self-Driven Vehicles, *TRANS. RES. PART A* 1, 12, acessível em https://www.researchgate.net/publication/303174405_Preparing_a_nation_for_autonomous_vehicles_opportunities_barriers_and_policy_recommendations_for_capitalizing_on_self-driven_vehicles.
- ❖ Ferreira e Dias Pereira (2017), Partilhar o Mundo com Robôs Autônomos: a Responsabilidade Civil Extracontratual por danos. Introdução ao problema, *Cuestiones de Interés Jurídico*, ISSN 2549-8402.
- ❖ Figueiredo Dias (2012), *Direito Penal – Parte Geral – Tomo I – Questões fundamentais; A Doutrina Geral do Crime*, Coimbra, Coimbra Editora.
- ❖ Fischer, J. M. & Ravizza, M. (1999), *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge University Press, ISBN 9780511814594 9780521775793.
- ❖ Flavell, J. H. (1979), *Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry*, *American Psychologist*, 34(10), 906-911, DOI: <http://dx.doi.org/10.1037/0003-066X.34.10.906>.
- ❖ Fleming, S. M., Dolan, R. J. e Frith, C.D. (2012), Metacognition: computation, biology and function, *Phil. Trans. R. Soc. B* 2012 367, DOI: 10.1098/rstb.2012.0021.
- ❖ Floridi, L. & Sanders, J. (2004), On the Morality of Artificial Agents, *Minds and Machines*, 14: 349, DOI: <https://doi.org/10.1023/B:MIND.0000035461.63578.9d>.
- ❖ Fraichard, T. (2014), Will the driver seat ever be empty? Report funded by Inria. <http://hal.inria.fr/hal-00965176>, acessado a 8 de Agosto de 2018.
- ❖ Frankfurt, H., 1971, Freedom of the Will and the Concept of a Person, *Journal of Philosophy*, 68(1): 5–20, obtido através de <http://www.sci.brooklyn.cuny.edu/~schopra/Persons/Frankfurt.pdf>.
- ❖ Garza, Andrew P. (2011), Look Ma, No Hands: Wrinkles and Wrecks in the Age of Autonomous Vehicles, *New Eng. L. Rev.* 46:58, obtido através de https://www.researchgate.net/publication/266223185_Look_Ma_No_Hands'_Wrinkles_and_Wrecks_in_the_Age_of_Autonomous_Vehicles.
- ❖ Genesereth, M.R., e Ketchpel, S.P. (1994), Software Agents, *Communications of the ACM CACM*, Volume 37 Issue 7, Pages 48-ff. DOI: 10.1145/176789.176794
- ❖ Goldman, A., (1970), *A Theory of Human Action*, Englewood Cliffs, NJ: Prentice-Hall, DOI: 10.2307/2217499.
- ❖ Goodall N.J., (2014), Ethical decision making during automated vehicle crashes, *Transp Res Rec J Transp Res Board* 2424:58–65, DOI: 10.3141/2424-07.

- ❖ Gogoll, J., Muller, J. F. (2017), Autonomous Cars: In Favor of a Mandatory Ethics Setting, *Sci Eng Ethics*, DOI 10.1007/s11948-016-9806-x
- ❖ Gray, J. (1921), *Nature and Sources of the Law*, acessível em <https://archive.org/details/in.ernet.dli.2015.507238/page/n3>, acedido no dia 16 de Setembro de 2018.
- ❖ Grimmelmann, J., *Regulation by Software* (2005). 114 Yale L.J. 1719 (2005). Available at SSRN: <https://ssrn.com/abstract=2358693>
- ❖ Gurbin, T. (2015), *Metacognition and Technology Adoption: Exploring Influences*. *Procedia -Social and Behavioral Sciences*. 191. 1576-1582, DOI: 10.1016/j.sbspro.2015.04.608.
- ❖ Hicks, D. J. (2018), *The Safety of Autonomous Vehicles, Lessons from Philosophy of Science*, DOI: 10.1109/MTS.2018.2795123
- ❖ Hart, H. L. A. (1965), *The Morality of the Criminal Law*, Magnes Press Oxford University Press.
- ❖ Ilková, V. ; Ilka, A. (2017), *Legal Aspects of Autonomous Vehicles – an Overview*, *Proceedings of the 2017 21st International Conference on Process Control (PC)*, Štrbské, Pleso, Slovakia, June 6 – 9, pp. 428-433, acessível através de http://publications.lib.chalmers.se/records/fulltext/249781/local_249781.pdf.
- ❖ Levy, J. (2016), *No Need to Reinvent the Wheel: Why Existing Liability Law Does Not Need to Be Preemptively Altered to Cope with the Debut of the Driverless Car*, 9 J. Bus. Entrepreneurship & L. 355, disponível em <https://digitalcommons.pepperdine.edu/jbel/vol9/iss2/5>.
- ❖ LaBar, K. & Cabeza, R. (2006), *Cognitive neuroscience of emotional memory*, *Nat Reviews, Neuroscience*, 7: 54-64, DOI: 10.1038/nrn1825.
- ❖ Lin, P. (2016), *Why Ethics Matters for Autonomous Cars*, em Maurer M., Gerdes J., Lenz B., Winner H. (eds) *Autonomous Driving*. Springer, Berlin, Heidelberg DOI: https://doi.org/10.1007/978-3-662-48847-8_4.
- ❖ Lonita, S. (2017), *Autonomous Vehicles: from paradigms to Technology*, *IOP Conference Series: Materials Science and Engineering*, Volume 252, conference, obtido através de <http://iopscience.iop.org/article/10.1088/1757-899X/252/1/012098>.
- ❖ Lowe, E. J. (2000), *An Introduction to the Philosophy of Mind*, Cambridge University Press, Cambridge.
- ❖ Luís, J. (2014), *Teoria do Comportamento Planeado, Psicopatía e a Intenção de Reincidir em Crimes Contra a Propriedade*, Tese de Mestrado em Criminologia, Faculdade de Direito da Universidade do Porto.
- ❖ Moniz Pereira, L., Saptawijaya, A. (2016), *Programming Machine Ethics*, Springer, DOI: 10.1007/978-3-319-29354-7
- ❖ Kalra, N., Anderson, J. e Wachs, M. (2009), *Liability and Regulation of Autonomous Vehicle Technologies*, California Path Program, Institute of Transportation, University of California, acessível através de https://www.researchgate.net/publication/228931139_Liability_and_Regulation_of_Autonomous_Vehicle_Technologies.
- ❖ Katrakazas, C., Quddus, M. Chen, W., Deka, L. (2015), *Real-time motion planning methods for autonomous on-road driving: State-of-the-art and future research directions*, *Transportation Research Part C: Emerging Technologies*, Volume 60, Pages 416-442, <https://doi.org/10.1016/j.trc.2015.09.011>.
- ❖ King, D. (2018), *Putting the Reins on Autonomous Vehicle Liability: Why Horse Accidents Are the Best Common Law Analogy*, 19 N.C.J.L. & Tech. 127, disponível em http://ncjolt.org/wp-content/uploads/2018/01/King_Final.pdf.
- ❖ Kone Mamadou Tadiou, 'Introduction to Artificial Intelligence', <<http://futurehumanevolution.com/artificial-intelligence-future-human-evolution/introduction-to-artificial-intelligence>> accessed 19May 2014.
- ❖ Kosko, G (1986)., *Fuzzy Cognitive Maps*, *International Journal of Man-Machine Studies*, Volume 24 Issue 1, Pages 65 – 75, DOI: 10.1016/S0020-7373(86)80040-2.
- ❖ Kuhn, Deanna & David Dean, Jr. (2004), *Metacognition: A Bridge Between Cognitive Psychology and Educational Practice*, *Theory Into Practice*, 43. 268-273, DOI: 10.1207/s15430421tip4304_4.
- ❖ Kuki, V. A. J. (2016), *Revisiting legal personhood*, PhD Candidate, University of Cambridge Paper for Spanish-Finnish Seminar in Legal Theory, Joensuu, obtido através de <http://www.uef.fi/documents/300201/0/Kurki+-+Revisiting+legal+personhood.pdf/56e99525-ba38-4c05-8034-3505d52d84a0>, consultado em Outubro de 2018.
- ❖ Leenes, R. E. and Lucivero, F. (2014), *Laws on Robots, Laws by Robots, Laws in Robots: Regulating Robot Behaviour by Design*, *Law, Innovation and Technology* (2014) 6(2) LIT 194–222. DOI: <http://dx.doi.org/10.5235/17579961.6.2.193> (2014) 6(2) LIT 193–220.
- ❖ LeValley, D. (2013), *Comment, Autonomous Vehicle Liability-Application of Common Carrier Liability*, 36 SEATTLE U. L. REV. SUPRA 5, 9.
- ❖ Marsella, S., and Gratch J. (2002), "A Step Towards Irrationality: Using Emotion to Change Belief." 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems, Bologna, Italy.
- ❖ Martinez, M. E. (2006), *What is Metacognition?*, *Phi Delta Kappan*, 87(9), 696–699, DOI: <https://doi.org/10.1177/003172170608700916>.
- ❖ Matthias, A. (2004), *The responsibility gap: Ascribing responsibility for the actions of learning automata*, *Ethics Inf Technol*, 6: 175. <https://doi.org/10.1007/s10676-004-3422-1>.
- ❖ Mele, A.R. (1992), *Springs of Action: Understanding Intentional Behavior*, Oxford: Oxford University Press
- ❖ Mele, A.R. (2003), *Motivation and Agency*, Oxford: Oxford University Press.

- ❖ Metcalfe, Janet & Son, Lisa K. (2012) Anoetic, noetic, and auto-noetic metacognition, In Michael Beran, Johannes Brandl, Josef Perner & Joëlle Proust (eds.), *The Foundations of Metacognition*, Oxford University Press.
- ❖ Ministério da Saúde. Direção-Geral da Saúde, Programa Nacional de Prevenção de Acidentes, Lisboa: DGS, 2010, 54 p.
- ❖ L. Minsky, Marvin. (1968), *Computation: Finite and Infinite Machines*, *The American Mathematical Monthly*, DOI: 75. 10.2307/2313471.
- ❖ MIT Autonomous Vehicle Technology Study: Large-Scale Deep Learning Based Analysis of Driver Behavior and Interaction with Automation, Lex Fridman*, Daniel E. Brown, Michael Glazer, William Angell, Spencer Dodd, Benedikt Jenik, Jack Terwilliger, Julia Kindelsberger, Li Ding, Sean Seaman, Hillary Abraham, Alea Mehler, Andrew Sipperley, Anthony Pettinato, Bobbie Seppelt, Linda Angell, Bruce Mehler, Bryan Reimer*, 2017.
- ❖ Moor, J. H. (1985), What is Computer Ethics, *Metaphilosophy*, 16: 266-275, DOI:10.1111/j.1467-9973.1985.tb00173.x.
- ❖ Morse, S. J. (2000), *Rationality and Responsibility*, Morse, Faculty Scholarship. Paper 524, obtido através de http://scholarship.law.upenn.edu/faculty_scholarship/524.
- ❖ Morse, S.J.. (2003), Diminished Rationality, Diminished Responsibility, *Ohio State Journal of Criminal Law*. 1. 289-308, obtido através de https://www.researchgate.net/publication/285715196_Diminished_Rationality_Diminished_Responsibility.
- ❖ MUÑOZ CONDE, F., ARÁN, M G. (1996). *Derecho Penal: Parte General*, 2ed, Valencia: Tirant Lo Blanch.
- ❖ Nelson, T. O. & Narens, L. (1990), Metamemory: a theoretical framework and new findings, *Psychol. Learn. Motiv. Adv. Res. Theory* 26, 125 – 173, DOI: 10.1016/S0079-7421(08)60053-5.
- ❖ Newell, A. (1994). *Unified Theories of Cognition*, Harvard University Press; Reprint edition, ISBN 0-674-92101-1.
- ❖ Nourbakhsh, R. Illah & Thorpe, E., Chuck. (1997), A Hybrid Human-Computer Autonomous Vehicle Architecture, Third ECPD International Conference on Advanced Robotics, Intelligent Automation and Control, obtido através de https://www.researchgate.net/publication/2516969_A_Hybrid_Human-Computer_Autonomous_Vehicle_Architecture.
- ❖ Shchetko, N. (2014), Laser Eyes Pose Price Hurdle for Driveless Cars, *Wall St. J*, consultado a 1 de Outubro.
- ❖ Solaiman, S.M. (2017), Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy, *Artif Intell Law*, 25: 155. DOI: <https://doi.org/10.1007/s10506-016-9192-3>
- ❖ Oliveira M., Santos V., Sappa A.D., Dias P. (2016) Scene Representations for Autonomous Driving: An Approach Based on Polygonal Primitives. In: Reis L., Moreira A., Lima P., Montano L., Muñoz-Martinez V. (eds) *Robot 2015: Second Iberian Robotics Conference. Advances in Intelligent Systems and Computing*, vol 417. Springer, Cham, DOI: https://doi.org/10.1007/978-3-319-27146-0_39.
- ❖ Olsaretti, S. (1998), Freedom, force and choice: Against the rights-based definition of voluntariness. *Journal of Political Philosophy* 6 (1):53–78, DOI: 10.1111/1467-9760.00046.
- ❖ Paden, B. & Čáp, Michal & Yong, Sze Zheng & Yershov, Dmitry & Frazzoli, Emilio (2016), A Survey of Motion Planning and Control Techniques for Self-driving Urban Vehicles, *IEEE Transactions on Intelligent Vehicles*. DOI: 1. 10.1109/TIV.2016.2578706.
- ❖ Palmerini, E., Bertolini, A., Battaglia, F., Koops, B. J., Carnevale, A., Salvini, P. (2016), RoboLaw: Towards a European framework for robotics regulation, *Robotics and Autonomous Systems*, Volume 86, Pages 78-85, DOI: <https://doi.org/10.1016/j.robot.2016.08.026>.
- ❖ Peacocke, C. (1979), Deviant Causal Chains. *Midwest Studies In Philosophy*, 4: 123-155. DOI:10.1111/j.1475-4975.1979.tb00375.x.
- ❖ Piaget J. (1976). *The Grasp of Consciousness*. Cambridge, MA: Harvard University Press
- ❖ Pietrzykowski, Tomasz. (2017), The Idea of Non-personal Subjects of Law, DOI: 10.1007/978-3-319-53462-6_4.
- ❖ Pizarro de Almeida, *Modelos de Inimputabilidade: da Teoria à Prática*, 1ª edição, Almedina, Janeiro de 2000.
- ❖ Pendleton, S.D.; Andersen, H.; Du, X.; Shen, X.; Meghjani, M.; Eng, Y.H.; Rus, D.; Ang, M.H.. (2017), Perception, Planning, Control, and Coordination for Autonomous Vehicles. *Machines* 5, 6, DOI: <https://doi.org/10.3390/machines5010006>.
- ❖ Petros Georgiades (2004), From the general to the situated: three decades of metacognition, *International Journal of Science Education*, 26:3, 365-383, DOI: 10.1080/0950069032000119401
- ❖ Resnick, M. (1994), Changing the Centralized Mind, *Technology Review*, pp. 32-40, obtido através de <https://ilk.media.mit.edu/papers/archive/CentralizedMind.html>.
- ❖ Ronald J. Brachman, Hector J. Levesque (2003), *Knowledge Representation and Reasoning* obtido através de <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.301.439&rep=rep1&type=pdf>.

- ❖ Russell, Norving (n 11) 234e321; Randall Davis, Howard Shrobe, Peter Szolovits, 'What is a Knowledge Representation?' (1993) *AI Magazine*, 14 (1):17e33 <<http://groups.csail.mit.edu/medg/ftp/psz/k-rep.html>>accessed 23.05.14.
- ❖ Ryan Whitman, How Google's Self-Driving Cars Detect and Avoid Obstacles, *ExtremeTech*, consultado a 1 de Outubro.
- ❖ Sack, D., Burgard, W. (2014), A comparison of methods for line extraction from range data, In *Proceedings of the 5th IFAC Symposium on Intelligent Autonomous Vehicles (IAV)*, Lisbon, Portugal, 5–7 July 2014; Volume 33.
- ❖ Salmond, J. W. (1913), *Jurisprudence*. 4th ed. London: Stevens and Haynes, obtido através de <https://archive.org/details/jurisprudence00salm>.
- ❖ Sandberg, A., & Bradshaw-Martin, H. (2013). In J. Romportl, et al. (Eds.), *What do cars think of trolley problems: Ethics for autonomous cars? . Beyond AI: Artificial Golem Intelligence*, Conference Proceedings. https://www.beyondai.zcu.cz/files/BAI2013_proceedings.pdf.
- ❖ Saxe R., Offen S. 2010 Seeing ourselves: what vision can teach us about metacognition. In *Metacognition and severe adult mental disorders* (eds Dimaggio G., Lysaker P. H.), pp. 13–30. Hove, East Sussex: Routledge.
- ❖ Schmidhuber, J. (2015), Deep learning in neural networks: An overview, *Neural Networks*, Volume 61, Pages 85-117, DOI: <https://doi.org/10.1016/j.neunet.2014.09.003>.
- ❖ Schraw, Gregory & Moshman, David (1995), *Metacognitive Theories*, *Educational Psychology Review*, 7. 351-371, DOI: 10.1007/BF02212307.
- ❖ Schlosser, Markus E. (2010). Bending it like beckham: Movement, control and deviant causal chains. *_Analysis_* 70 (2):299-303, DOI: 10.1093/analys/anp176.
- ❖ Searle, J.R., 1983, *Intentionality: An Essay in the Philosophy of Mind*, Cambridge: Cambridge University Press. Searle, 1999.
- ❖ Shardlow, N. (1990), *Action and agency in cognitive science*, Master's thesis, Department of Psychology, University of Manchester, Oxford Rd., Manchester M13 9PL, UK.
- ❖ Shoham, Y. (1993), *Agent-Oriented Programming*, *Artificial Intelligence*, Volume 60, Issue 1, pp. 51-92, DOI: 10.1016/0004-3702(93)90034-9
- ❖ Silber, John R. (1967), "Being and Doing: A Study of Status Responsibility and Voluntary Responsibility," *University of Chicago Law Review*: Vol. 35 : Iss. 1 , Article 4., acessível em <https://chicagounbound.uchicago.edu/uclrev/vol35/iss1/4>
- ❖ Taylor, R., 1966, *Action and Purpose*, Englewood Cliffs: Prentice-Hall, DOI: 10.2307/2024334.
- ❖ Pagallo, U. (2013), *The Laws of Robots - Crimes, Contracts, and Torts*, Springer, DOI: 10.1007/978-94-007-6564-1.
- ❖ C. Vladeck, David. (2014). *Machines without principals: Liability rules and artificial intelligence*. *Washington Law Review*. 89. 117-150, obtido através de https://www.researchgate.net/publication/288714455_Machines_without_principals_Liability_rules_and_artificial_intelligence.
- ❖ Welty, C. A. (1996), *An Integrated Representation for Software Development and Discovery*, Disertation, obtido através de <<http://www.cs.vassar.edu/~weltyc/papers/phd/HTML/dissertation-1.html>> accessed 23.05.14.
- ❖ Wilks, Y.(1985), *Machine Translation and AI: Issues and their Histories (MCCS – 85-29)*, Las cruces – Computing Research Laboratory, New Mexico State University, acessível através de https://www.researchgate.net/publication/241488724_MACHINE_TRANSLATION_AND_ARTIFICIAL_INTELLIGENCE.
- ❖ Wooldridge e Jennings (1995), *Intelligent Agents: Theory and Practice*, *The Knowledge Engineering Review*, Volume 10, Issue 2, pp. 115-152, DOI: <https://doi.org/10.1017/S0269888900008122>.
- ❖ Yaffe, G. (2012), *The Voluntary Act Requirement*, Em Marmor Andrei (ed), *The Routledge Companion to Philosophy of Law*, Routledge, pp. 174, obtido através de <https://philpapers.org/rec/YAFTVA>.
- ❖ Zohn, J. R. (2015), *When Robts Attack: How Should the Law Handle Self-Driving Cars That Cause Damages?* *University of Illinois Journal of Law Technology and Policy*.